FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Automatic Assessment of Infantile Hemangiomas

Pedro Guilherme Reis Alves

Mestrado Integrado em Bioengenharia

Supervisor: Jaime dos Santos Cardoso, PhD

July 14, 2017

Resumo

Os Hemangiomas Infantis correspondem ao tipo de tumor vascular benigno mais comum que afecta crianças. O seu crescimento pode durar vários meses até entrar na chamada fase de involução, onde diminui progressivamente de tamanho e de intensidade de cor. No entanto, o crescimento descontrolado e a localização da lesão podem colocar em risco certas estruturas vitais, como a visão ou as vias respiratórias, e o aparecimento de complicações, como a ulceração ou desfiguração. Um diagnóstico, monitorização e tratamento adequados e atempados podem ajudar a prevenir ou diminuir estas complicações. No entanto, não existe definido um protocolo de diagnóstico e acompanhamento da lesão que seja concreto e universal. A lesão é avaliada subjectivamente pelo médico, que estima a área através da medida da largura e comprimento da lesão, e estima o grau de regressão pela área do hemangioma que não apresentar a típica cor vermelha intensa. Para se obter medidas mais concretas e objectivas, é necessário um sistema de avaliação mais padronizado para este tipo de lesões. Um exemplo são os sistemas de Diagnóstico Assistido por Computador (DAC).

Neste tipo de sistemas, e em especial naqueles que se focam na avaliação da lesão contida na imagem que lhes é fornecida, a primeira etapa consiste na segmentação da área afectada. Contrariamente ao que acontece no campo do melanoma, a segmentação de imagens de hemangioma infantil é um tópico relativamente inexplorado, com apenas alguns trabalhos desenvolvidos nos últimos anos. Assim, o foco principal desta dissertação será o estudo de várias metodologias para a segmentação do hemangioma infantil. Os métodos estudados baseiam-se em metodologias apelidadas de "superficiais" - método de Otsu, algoritmo de Canny, fuzzy *c*-means, GVF snakes e o cálculo do caminho mais curto para gerar um contorno fechado - e metodologias "profundas" - redes neuronais convolucionais.

Para avaliação da performance destes métodos foi criado um dataset a partir de fotografias das lesões de vários pacientes, tiradas em várias consultas, de uma maneira não controlada, ou seja, sem preocupação em relação à posição da câmara, distância à lesão ou iluminação do espaço. Estas fotografias foram adquiridas com autorização prévia dos pais das crianças.

Uma etapa de pré-processamento foi desenvolvida para facilitar o processo de segmentação. A ideia foi eliminar a variabilidade existente entre pacientes e entre imagens do mesmo paciente. Para isto foram implementadas técnicas de uniformização de cor e de iluminação entre as imagens e de eliminação de ruído.

De entre os métodos estudados, os que obtiveram melhores resultados foram o método do caminho mais curto (93.37% de accuracy, 90.35% coeficiente de Dice e 83.04% de índice de Jaccard) e uma rede neuronal convolucional que recebe duas imagens com resolução diferentes, centradas no mesmo pixel (92.99% accuracy, 90.20% coeficiente de Dice e 82.95% de índice de Jaccard). Apesar de muito semelhantes, o método do caminho mais curto revelou-se o melhor, pois é o que consegue consegue a melhor aproximação entre a borda segmentada e a real.

Com base nestes resultados, conclui-se que a informação acerca do contorno da lesão se revela muito importante, e uma característica mais discriminativa quando comparada com a cor do

hemangioma. Isto deve-se essencialmente ao facto da cor da lesão ser muito variável de paciente para paciente e de se alterar muito aquando da fase de involução.

Foi também feito Transfer Learning, partindo de uma rede convolucional treinada em melanoma que é depois refinada para segmentar hemangiomas. Neste trabalho foi feita transferência entre métodos de aquisição diferentes - de dermoscopia para smartphone (87.94% de accuracy, 83.30% coeficiente de Dice e 73.05% de índice de Jaccard) - e entre domínios, mas com o mesmo sensor - de melanoma para hemangioma usando smartphone (87.84% de accuracy, 83.47% coeficiente de Dice e 73.14% de índice de Jaccard). Os resultados mostram que esta é uma área promissora e que o refinar de soluções pode poupar tempo e recursos quando o número de dados não é suficiente para uma aplicação específica. No entanto, permanece uma tarefa díficil, dadas as diferenças entre melanoma e hemangioma infantil.

Assim, as conclusões principais desta tese são: apesar da informação da cor e da borda, de forma independente, proporcionarem bons resultados, a sua fusão poderia aumentar a performance das metodologias testadas; as redes neuronais convolucionais constituem uma ferramenta poderosa pelo seu poder de aprendizagem e generalização, mas outro tipo de arquitecturas mais complexas ou mais fontes de informação poderiam melhorar os resultados; a transferência de aprendizagem de tarefas semelhantes revela-se um tópico interessante que merece ser explorado em trabalhos futuros.

Abstract

Infantile Hemangiomas make up for the most common type of benign vascular tumors affecting children. They can grow for several months until beginning to involute. However, their growth can be too aggressive or located in sensitive areas, which can result in the impairment of vital structures or the arising of complications, such as ulceration and disfigurement. An early diagnosis and adequate treatment at the right time can prevent or lessen such complications. However, in present-day clinical practice there's no objective standard diagnosis and monitoring protocol. The lesion is visually assessed and an estimation of its features is done by the clinician, being subjective and varying between practitioners. The lesion's area is estimated by width and height measurement and the degree of regression is estimated by the amount of graying of the lesion's reddish color. For more objective measures, an automatic evaluation system (CAD system) is needed to aid clinicians in assessing with more certainty these measures and the effectiveness of a given patient's response to a treatment.

In this kind of systems the first stage consists in segmenting the affected area. Contrary to melanoma, the segmentation of IH lesions has not been a subject of much focus. Therefore, the core of this thesis will be the evaluation of various methodologies for the segmentation of the IH. The methods addressed were divided in "shallow" - Otsu's method, Canny's algorithm, fuzzy c-means clustering, GVF snakes and shortest path calculation to obtain a closed contour - and "deep" methodologies - convolutional neural networks.

To evaluate the performance of such approaches, a dataset was created from photos of IH lesions of several patients from various consultations, acquired in a unconstrained manner, meaning that there were no regard for camera position, distance or room lighting. These photos were acquired with the parent's previous consent.

A preprocessing stage was developed to aid in the segmentation process. The idea was to mitigate the existing variability between images. To this end were implemented techniques for color constancy, illumination equalization and noise removal.

From the tested methods, the ones that obtained the best results were the computation of the shortest path (93.37% accuracy, 90.35% Dice coefficient and 83.04% Jaccard index) and a convolutional neural network that has two different inputs (92.99% accuracy, 90.20% Dice coefficient and 82.95% Jaccard index). Despite very similar, the shortest path method revealed itself the best, as it is the one that can provide a better approximation between the segmented border and the real one.

With these results, one recognized that information about the edges of the lesion are an important and more discriminate feature than when compared with the color of the lesion. This is mainly due to the fact that color presents a high degree of variability between patients and that it changes completely when the lesion starts to involute.

To explore additional approaches, Transfer Learning was also employed using a convolutional neural network pre-trained on melanoma images. Learning was transferred in two different ways: from different acquisition sensors - from dermoscopy to smartphone camera (87.94% accuracy,

83.30% Dice coefficient and 73.05% Jaccard index) - and from different domains with the same acquisition sensor - from melanoma to hemangioma using a smartphone (87.84% accuracy, 83.47% Dice coefficient and 73.14% Jaccard index). Results show that this is a promising area and that the approach of refining existing solutions for similar tasks can still provide good results while saving time and resources when the volume of data is not enough for a given application. However, it is still a difficult task, given the differences between melanoma and IH.

Thus, the main conclusions drawn from this thesis were: despite the fact that IH color and border information can, independently, achieve good performances, their fusion could help enhance the obtained results; the tested CNNs proved to be a powerful tool, given their generalization and learning ability, but more complex network architectures or more sources of information, could help improve segmentation; transfer learning has proven to be an interesting and promising topic that deserves to be better explored in future works.

Acknowledgments

Em primeiro lugar, uma forte palavra de agradecimento ao Prof. Jaime Cardoso pela sua disponibilidade, orientação e ideias ao longo do desenvolvimento desta tese. Mais uma vez revelou-se um mentor fantástico e alguém que levo como exemplo, tanto a nível professional como pessoal.

Ao Ricardo Cruz, pela ajuda sempre que necessitei e que sempre se mostrou disponível.

Agradeço ao grupo do Centro Hospitalar do São João pelo fornecimento dos dados, sem os quais esta dissertação não seria possível.

Aos meus pais, por tudo, sempre.

Há 5 anos passei a fazer parte de um grupo que, aos poucos e poucos, e sem me aperceber, me faria redefinir vários princípios e valores que antes tomava como certos. No meio dessas pessoas e ao lado de quem começaria esta jornada comigo, vivi momentos únicos e indescritíveis, que agora me apercebo que foram um privilégio. Sem haver necessidade de cair no cliché de referir este ou aquele, farei o que aprendi e dirijo, não um obrigado, mas uma palavra de apreço a todos os que me ensinaram e inspiraram e a todos os que, espero eu, tenha conseguido ensinar e inspirar.

Pelas noites mal dormidas, pelos dias cansados, pelas pausas de café, pelas conversas longas, por todas as gargalhadas e por todas as lágrimas, pelas aventuras e pelas histórias que hoje tenho para contar. Que todos os que sabem que este texto lhes diz respeito me permitam o egoísmo de vos continuar a ter ao meu lado sempre que precisar. Eu certamente estarei ao vosso.

Valete, Frates.

Pedro G. Alves

Cheio de Deus, não temo o que virá,
Pois, venha o que vier, nunca será
Maior do que a minha alma.

Fernando Pessoa, *Mensagem*

Contents

List of Figures

List of Tables

Abreviaturas e Símbolos

CAD	Computer-Aided Diagnostic
CNN	Convolutional Neural Network
FCM	Fuzzy c-Means
GVF	Gradient Vector Flow
IH	Infantile Hemangioma
ReLU	Rectified Linear Units
ROI	Region Of Interest
SP	Shortest Path

Chapter 1

Introduction

The present thesis refers to the relatively unexplored topic of computer-assisted monitoring and evaluation of Infantile Hemangiomas. They represent the most common type of vascular tumor affecting children and can pose a threat to vital structures. In this section, we start by motivating the need of further research and needed improvement in this topic. Then, we present the main goals and contributions of this work. Finally, the strcuture of the remaining document is outlined.

1.1 Motivation

Infantile Hemangiomas (IH) are the most frequent benign vascular tumors in infancy, with an incidence around 4% to 5% (Kilcline and Frieden, 2008) and higher prevalence in females and Caucasians (Chiller et al., 2002; Dickison et al., 2011). They go through growth and involution phases and in most cases are apparent during the first weeks of life. For many patients the involution phase ends without significant sequelae, but IH frequently occur in areas that present a risk due to the potential for rapid proliferation. With indiscriminately growth they can threaten structures by compression or obstruction and hinder vital functions like vision, breathing and ingestion (Chiller et al., 2002; Haggstrom et al., 2006). Also, they present a risk for permanent disfigurement when appearing in certain areas that are cosmetically sensitive, like the face and neck. To avoid these consequences and induce regression, an appropriate treatment, dependent on the type of lesion at hand, and at an appropriate time, needs to be established.

Current clinical practice does not dictate a standard diagnosis protocol for infantile hemangiomas nor a reliable way for predicting its evolution. The hemangioma area is approximated by the width's and height's measurement with a ruler by the clinician during examination, and the degree of regression is assessed by visual estimation of the percentage of graying of the lesion's reddish color (Mattassi et al., 2009; Püttgen, 2014; Zambanini et al., 2006, 2007).

These methods have low reliability, significant variability and are imprecise, as the hemangioma shape is not taken into consideration to calculate the affected area. The regression degree estimation depends on the experience and rating of the clinician in a subjective manner, which limits the discernment of the effectiveness of a certain treatment. Therefore, there is a need for an

evaluation system that could support practitioners with accurate measurements for detecting and quantifying changes in the hemangioma over time. Such objective measurements would help in the decision of whether to start treatment, to evaluate the application of different treatments in clinical studies, and to evaluate the response of an individual patient to a certain treatment (Zambanini et al., 2010).

Over the years, computer-aided diagnosis (CAD) systems have been a subject of great interest for the detection and analysis of skin lesions from digital images, as they can help dermatologists easily identify malignant lesions, a process that is very time consuming and of subjective interpretation (Celebi et al., 2007; Scharcanski and Celebi, 2014; Celebi et al., 2015). Particularly, they can help diagnosing malignant lesions at their initial stage, which is when the patient has higher probability of cure, can be used to monitor benign lesions and help prevent the development of malignancy, or help deciding if a certain therapy should be initiated (Oliveira et al., 2016).

Standardized image acquisition for lighting, color and position of the patient could improve the clinician efficiency in quantifying multiple characteristics over time. This would help overcome the lack of quantitative information about the differences in growth patterns for the various stages of IHs. Despite these advantages, having a defined acquisition protocol also has its drawbacks. We are talking about acquiring during a patient's consultation, something that is scheduled for a specific time and that should proceed as swiftly as possible. The addition of this acquisition stage would delay the appointment's normal duration. Another important question is that the patients in question are children, which raises some obstacles: parents may not be too willing on allowing their children to be part of studies, especially if includes photos being taken; infants are not as cooperative as adults and acceptable photos could be hard to acquire. Therefore, a controlled acquisition setup would pose some usability and practicality issues to which the pediatric environment may not yet be prepared to face.

Another aspect that would prove beneficial would be the use of multimodal imaging in the assessment of IHs, as it would make available different types of physical information. Specifically, the extraction of visual information with respect to both morphology and function (e.g. blood flow, temperature), which are then translated to quantitative values. This would provide useful information about a large number of significant features of the hemangioma, adding objectivity and reproducibility to the classification and treatment efficacy of IHs (Burkes et al., 2016). It has been well established that multimodality can increase the accuracy and completeness of information, given that each modality has its own limitations (Vogt and Ermert, 2006).

With this motivation in mind, this work will address the development of an automatic monitoring system for the quantitative assessment of IH. The first step in such a system is the segmentation of the region of interest, that is, the hemangioma; then, measurements regarding its area and regression can be computed. If follow-up images are available, the system should evaluate the evolution of the hemangioma throughout time. With this information, the system can help the clinician to have more certainty in his decision on whether or not to start treatment, which treatment would be most suitable, how is the hemangioma's response to it, its evolution over time, and finally, when to stop the treatment.

1.2 Research Goals and Contributions

The aim of this project is the development of a CAD system for the automatic monitoring of infantile hemangiomas. Such a system can be divided into 3 stages:

1. Hemangioma Risk Estimation
2. Hemangioma Segmentation
3. Hemangioma Evaluation
   (a) Extraction of lesion's measurements (width, height, area);
   (b) Registration of follow-up lesions
   (c) Regression detection and quantization

Given the time frame for the development of this thesis, the focus of this work was the second stage: segmenting the lesion. The main scientific contributions of this thesis are:

- Awareness and stimulation of research of the topic of IH, as there are very few works developed in the last decade in this area.
- Development of a framework for infantile hemangioma lesions segmentation. The developed system presented overall results superior to state-of-the-art approaches and showed to be robust to different acquisition settings and image resolution.
- Creation of a dataset from photos acquired in unconstrained settings, over the course of various consultations, that were manually annotated and prepared. This dataset can be used to develop algorithms that are robust to environment changes and develop methodologies to measure the degree of involution over time and assess the efficacy of a given treatment.
- Study on transfer learning between melanoma and infantile hemangioma tasks.

In the next section, the structure of this dissertation is outlined.

1.3 Dissertation Structure

The remaining of this document is organized in 5 chapters. The motivation and introduction to the current problems concerning IHs were already covered in this first chapter.

Chapter 2 will present a more detailed overview on infantile hemangiomas followed by a review on the state-of-the-art approaches for this topic.

Chapter 3 will focus on the theoretical background behind the segmentation and machine learning methodologies used in the development of the present work.

In Chapter 4 the obtained results from the various approaches are shown and discussed and finally, in Chapter 5, the main conclusions are presented, as well as the future lines of research regarding the topics that were explored throughout this dissertation.

Chapter 2

Infantile Hemangiomas: a Literature Review

In this chapter, the issue of infantile hemangiomas is addressed, covering its physiology, epidemiology, associated complications and diagnostic assessment. The problem associated with current clinical practices is explained, reinforcing the need for more research and guidelines. A review of the state-of-the-art in this subject is also presented, exploring the work that has been done in this area.

2.1 Overview on Infantile Hemangiomas

Infantile Hemangiomas (IH) are the most common vascular tumors of infancy, being characterized by rapid growth followed by spontaneous and slow involution. The mechanisms that contribute to its appearance and proliferation are still poorly understood. It is likely that the origin of IH is multifactorial, with genetic factors being part of the trigger, but main causes are attributed to vasculogenesis and angiogenesis (Greenberger and Bischoff, 2013).

Vasculogenesis is the *de novo* formation of blood vessels from stem cells and angiogenesis is the growing of blood vessels from pre-existing ones. Angiogenesis occurs primarily during embryogenesis and postnatal life, and it is a complex process that requires many different steps involving extensive interplay between cells, soluble factors and extracellular matrix components. Due to being such a complex process, it is prone to failure in the regulation of cell proliferation.

Vascular anomalies can be mainly divided in malformations and tumors. Vascular malformations result from aberrant development of the vascular elements during embryogenesis and fetal maturation. Vascular tumors occur when the balance between cell proliferation and apoptosis is disrupted and there is a lack of angiogenic inhibitors, causing the uncontrolled proliferation of endothelial cells of blood vessels (Mattassi et al., 2009).

The most frequent benign vascular tumor in infancy is the infantile hemangioma (Figure 2.1), with an incidence of around 4% to 5% (Kilcline and Frieden, 2008). It is more frequent in Caucasians, females and premature children, especially those with low weight (Dickison et al., 2011;

Chiller et al., 2002; Tollefson and Frieden, 2012). Generally, they exhibit three main evolutive phases: a *proliferative phase*, where the tumor grows rapidly, a *maturation/plateau phase*, where it ceases to grow, and an *involution phase*, in which the color of the lesion is attenuated and its size diminishes (Mulliken and Glowacki, 1982; Esterly, 1995; Boon et al., 1996). The duration of each phase is dependent on the type of IH, but most reach 80% of their final size by a mean age of 5 months (Chang et al., 2008), gradually involuting over the years, being complete in most cases by the age of 4 (Finn et al., 1983; Mulliken and Glowacki, 1982). An example of lesion involution is showed in Figure 2.1.

Figure 2.1: A typical infantile hemangioma lesion (left) and the same lesion in its involution stage (right).

Based on clinical features, hemangiomas can be classified as *segmental*, *localized*, *indeterminate* and *multifocal* (Figure 2.2). Segmental hemangiomas (2.2a) are those with a plaque-like morphology and geographic patterning over a specific cutaneous region; localized hemangiomas (2.2b) are those that grow from a single focal point; indeterminate hemangiomas (2.2c) are those that cannot be readily classified as segmental or localized; lastly, multifocal hemangiomas (2.2d) are those that have eight or more cutaneous lesions present (Chiller et al., 2002). Haggstrom et al. (2006), Chiller et al. (2002) and Tollefson and Frieden (2012) found in their prospective studies that localized hemangiomas are the most predominant subtype, followed by segmental hemangiomas, and that indeterminate and multifocal hemangiomas are the least predominant subtypes, respectively. Another important clinical feature is the location of the IH, which appears predominantly on the face, followed by head and neck, trunk, extremities and perineum (Tollefson and Frieden, 2012; Dickison et al., 2011; Haggstrom et al., 2006).

While being benign, many IH have various related complications which include ulceration, functional difficulties, airway obstruction, disfigurement and scarring, auditory canal occlusion and vision impairment (Chiller et al., 2002; Haggstrom et al., 2006). Ulceration is the most frequent and can lead to pain, bleeding and infection, mainly affecting the perineal and perioral regions (Haggstrom et al., 2006; Liang and Frieden, 2014; Püttgen, 2014; Chiller et al., 2002). However, the characteristics that are mostly associated with the development of complications are big size, segmental subtype and being located on the face (Haggstrom et al., 2006; Chiller et al., 2002; Frieden et al., 2005).

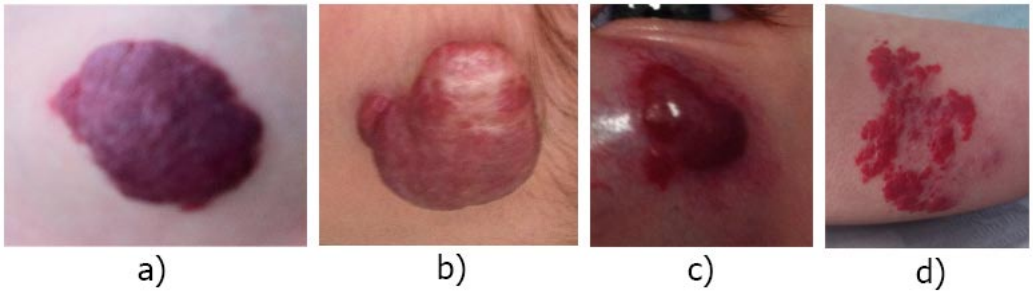

Figure 2.2: Examples of IH subtypes: segmental (a), localized (b), indeterminate (c) and multifocal (d).

To prevent these complications, an adequate diagnosis must be delivered to determine if treatment is needed to stop further proliferation and induce regression, or to avoid scarring and disfigurement (Haggstrom et al., 2006; Dickison et al., 2011; Liang and Frieden, 2014). Given that hemangiomas are so heterogeneous, deciding if treatment is needed and when it is needed requires detailed knowledge of the IH natural history and must be made individually for each patient (Püttgen, 2014). Also, if risk for disfigurement has been established, timing is one of the most important factors to decide whether to intervene (Luu and Frieden, 2013).

Since hemangiomas may spontaneously start to involute, waiting for this regression remains a viable treatment option. However, this principle is wrong, unless it serves as a control for cases such as small, uncomplicated IHs in areas that do not pose a problem and without any tendency to proliferate. If growth has not been excluded, then this control has to be frequent. Even for regressing hemangiomas, this waiting principle should only be applied if there's no risk of complication involved (e.g. ulceration). Undergoing a certain therapy may cause adverse local or systemic side effects, particularly disfigurement and scarring, and thus intervention has been reserved for patients with significant complications. Therefore, for the so called "small problem" hemangiomas, it is difficult to choose a therapy that can eliminate the lesions before the arise of complications and without side effects. The aim of therapy for IHs is not the immediate removal, but rather to stop proliferation, speed up regression and avoid/remove any existing functional problems (Mattassi et al., 2009; Luu and Frieden, 2013).

Current clinical practice consists in the examination of the IH features, such as color, shape, size and subtype, to determine if there is any risk for proliferation or if it has already started involuting. The hemangioma's height and width is manually measured and the affected area estimated, while the regression degree is visually estimated by the amount of fading (graying) of the normal saturated red color (Mattassi et al., 2009; Püttgen, 2014). Naturally, this is an inaccurate and subjective approach since that the ratio *affected area* / *area reduction* is very difficult to measure. This leads to strongly different evaluations between physicians, with the estimation of the degree of regression being based on the classification of a single individual, that may or may not have experience with such cases (Chang et al., 2008; Zambanini et al., 2010). As a result, hemangiomas that should undergo therapy may not receive it because of the subjective assessment. Therefore, an automatic method to quantitatively assess cutaneous hemangiomas is needed to provide prac-

titioners with precise, accurate and faster measurements than those taken by manual procedures, that can help decide whether treatment is suitable or not. This would allow for the measurement of the area to become clinically relevant and would help improve the quality and conclusions of clinical trials.

In the next section, the state-of-the-art regarding the outlined problem is reviewed.

2.2 State-of-the-Art Review

Previous to 2006, there are no reported work of systems to automatically assess images of hemangioma lesions (Zambanini et al., 2006). Nonetheless, skin lesions like melanoma or psoriasis have been thoroughly researched. Oliveira et al. (2016) present an excellent review of the state-of-the-art for the segmentation of pigmented skin lesions, but where IH are not included. They show that thresholding methods are the most widely used, due to their simplicity, overall good performance and computational efficiency. Artificial Intelligence based techniques are also widely used due to the advantages of being able to learn from the data and optimized for the "best" segmentation result. Other frequently proposed algorithms are those based on active contour models, like the well-known Snakes, and those based on region growing and merging.

Zambanini et al. (2006) were the firsts to develop a method to segment the hemangioma and measure its area. They argue that methods used in other types of skin lesions cannot be applied due to the low contrast between the skin and hemangioma. Since hemangiomas appear in various shapes and sizes, making use of specific a priori knowledge about its general appearance is difficult. Therefore, they performed segmentation based on color features using a single layer perceptron that classifies the pixels as hemangioma or non-hemangioma. First, a median filter is applied and the areas that contain no skin are masked out. The images are normalized in respect to the skin color, in order to achieve a more accurate classification result. Then, images are converted to the HSV color space, from which the H channel is retained, and converted to the CIE L*a*b* color space from which the $a*$ channel and the weighted Euclidean distance between the $a*$ and $b*$ channels are retained. These three features along with the G channel from the RGB color space constitute the set of features that are used in the classifier. Results were obtained on a set of 120 images taken with a digital camera, from which 30 were used as training set and 90 were used as test set. They obtained a mean accuracy of 93.2% and a border error of 32.1%. There is no information on how the acquisition of the photos was carried out. The same authors continued their research (Zambanini et al., 2010) on this problem and developed a method for the assessment of the lesion's evolution from images taken from follow-up examinations. Segmentation was performed using the same algorithm described above followed by registration. Then, detection of the regions exhibiting regression is made based on color features, using a pixel-wise classification with a multilayer perceptron. This is due to the fact that, unlike methods used in other types of skin lesions, single measurement of the entire affected area is not enough to have a complete assessment of regression, as the process starts with color changes and not partial disappearance of the lesion area.

More recently, Sultana et al. (2015b) preprocessed the image with the contrast-limited adaptative histogram equalization (CLAHE) technique for contrast enhancement, a Nagao filter for noise removal and an alpha blending approach to eliminate variations in illumination. Segmentation was performed using a two-step thresholding in the a^* channel of the CIE L*a*b* color space. Experiments were performed on a set of 25 images taken with a digital camera in which the acquisition process was not rigorous concerning lighting, camera distance and angle, or hemangioma positioning. These are true-color (24 bits per pixel) and have a resolution of 4324×3240 pixels. The authors do not disclose any quantitatively performance results, but state that the border error is calculated. In another work (Sultana et al., 2015a), the authors use a color constancy approach to preprocess the images and correct illumination. Color constancy is used to remove the effects that the color of a light source could have on an image. They devised an algorithm that chooses the best out of five existing color constancy transformations, for each case independently. This process is done by applying all five methods to two subsequent hemangioma images from the same patient and computing the Euclidean distance between the average skin color of both images. The best color transformation is the one with the minimum calculated distance. The rationale is that the only aspect that remains constant between successive acquisitions is the patient's skin color. This algorithm is then applied as part of the preprocessing step. Segmentation of the lesion area was performed like in their previous work, on the same set of images, on which they report a detection rate of 85% and a false positive rate of 16.4%.

Oprisescu et al. (2015) concluded in their work that the best color space to use for segmentation is the CIE L*a*b* space. They propose a segmentation procedure based on the following steps: construction of a binary edge density image using the Canny edge detector and a blurring kernel; building of the hemangioma learning image taken from inspecting the R component of the RGB space of this previous image; computing the mean and covariance matrices for Maximum a Posteriori (MAP); finally, performing MAP-Markov segmentation. The idea is to auto-determine the learning sets for both classes, which are then used to compute the matrices for MAP segmentation. Since the hemangioma is usually very textured, the edge detection will reveal a higher contour density in the hemangioma area. Inspecting the red component of the pixels with high contour density will result in the hemangioma learning set. A regularization phase using discrete Markov fields is applied to improve the obtained segmentation results. They used a dataset of 40 RGB images of size 800×600 pixels acquired with a digital camera at the Marie Curie children's Hospital from Bucharest, on which they obtained a border error of 19.49%.

Neghină et al. (2016) first convert the images to the CIE L*a*b* color space and retain the a^* channel. To this image, a contrast stretching is performed, so that the pixel's intensity values lie between 0 and 255. This preprocessed image is used as input for all the used segmentation methods: Otsu's method, Fuzzy c-Means clustering (FCM) and Region Growing based on FCM. They used a set of 33 images, taken in different conditions of illumination, with different cameras, and with the child standing naturally in different positions. The best obtained result was with the Region Growing FCM based method, on which they calculated the average between the number of true positives and true negatives, that resulted in a score of 90.27%. In another work (Neghină

et al., 2016), the authors apply the same preprocessing, but use a segmentation method based on Self Organizing Maps followed by a Morphological Method of Reducing the Number of Classes. The final step consists in computing the smoothness and then determine the mask with the cleanest edges. The results are obtained for the same dataset, with a reported performance increase of 1.06% relatively to the result of the previous work using Region Growing FCM.

From the presented state-of-the-art it can be seen that some limitations exist and that there is room for improvement. First, in most works, the preprocessing steps are scarce, like in the work by Oprisescu et al. (2015) and Neghină et al. (2016). Proper preprocessing of the images could greatly improve the subsequent steps, given that the images often contain noise, heterogeneous illumination, low contrast and different resolution of the lesion between images.

In the segmentation step, only Neghină et al. (2016) compare various methods and in all works only color features are used to segment the hemangioma. The use of more features related to shape and texture, as well as the ensemble of various methods, could help improve segmentation and deal with more troublesome images. One limitation common to all works is the use of standard digital color images to assess IHs. The use of more sources of information, such as infrared images, images from thermography, or other skin imaging techniques could help extract more relevant information to aid the clinician's decision. Therefore, the use of multimodality would be seen as an advantage to be explored.

Lastly, only the work of Zambanini et al. (2010) engages the use of follow-up images, which can be of great use to assess the degree of regression over time and evaluate the efficacy of a given therapy. From single-time sets of images, it can only be assessed the stage and degree of regression at that moment. Nonetheless, it its noted that the main goal of most of these works is the development of better preprocessing methods or better segmentation solutions.

Given all the different aspects inherent to the different goals of this project, the state-of-the-art methods can serve as a basis for the development of the CAD system, with the needed improvements. In the next chapter, the theorethical background of the various tools and methodologies used in the development of this work are thoroughly discussed.

Chapter 3

Computer-Assisted Monitoring of Hemangiomas

This chapter will start by addressing the database composition and the processing needed in the next stages, as well as cover the theoretical background concerning the methodologies used in the development of this work. These methodologies fall into three different stages: preprocessing, lesion segmentation and postprocessing. A brief overview of some advantages and disadvantages of each method is also given, that will be supported and discussed in Chapter 4.

3.1 Framework for Lesion Segmentation

As the main goal is the segmentation of the IH, a variety of methods were chosen in order to evaluate different approaches. First, some shallow methodologies were considered, like Otsu's method, active contours and clustering, based on reported work from the state-of-the-art. These methods employ different techniques to perform segmentation and rely on different features: Otsu's method and FCM clustering rely on pixel intensity, while snakes rely on the edges of the lesion. A deep learning approach was also considered, with the use of CNNs. The networks use a set of filters to obtain a certain response to the features of the image, that may be based on color or shape features of the input.

As with all segmentation methodologies, a way to assess its performance is needed. To this end, a database of images was created, a procedure that will be detailed in the next section.

To aid the segmentation stage and reduce variability in the dataset, a preprocessing stage was devised, consisting of a illumination equalization step, a color constancy step and an image smoothing step. Also, a common step in image processing that follows lesion segmentation is postprocessing the results, to try and eliminate misclassified pixels. All these steps are shown in the block diagram of Figure 3.1, and will be thoroughly detailed in the next subsections.

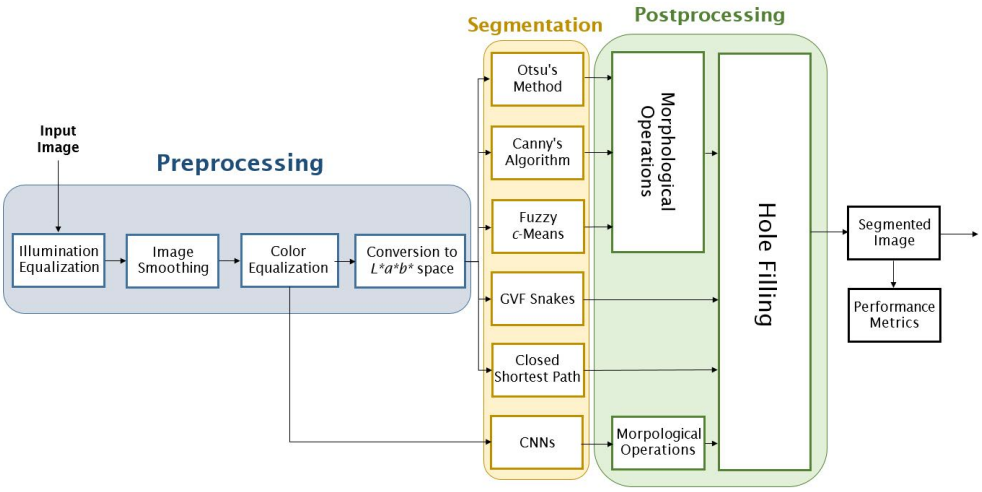

Figure 3.1: Block diagram for the proposed framework, showing the different methodologies in each major stage.

3.2 Database Composition

Before going into the detailed steps of the framework shown in the previous section, let us first address the dataset that was constructed to evaluate its performance.

This database is composed of photos taken with a smartphone during patient's consultation over several years. These were taken without any established acquisition protocol, meaning that no defined position, lighting, camera distance or any other constraints were taken into account. The main reasons for using these images in the development of this work are:

- **Availability:** given that appointments are weeks or months apart and that the number of patients is scarce, it would not have been possible to establish a good acquisition protocol and setting and acquire an adequate number of images in the given time frame. Therefore, as the doctor had a great number of photos taken over the years, the opportunity to use them arose.

- **More realistic scenario:** the photos composing the dataset were taken by the doctor with a smartphone with the only concern that the whole hemangioma or a specific part was visible. The lighting, camera or patient position and camera distance were not taken into account. Therefore, this constitutes a more realistic setting, that presents both advantages and disadvantages that will be discussed ahead.

A total of 200 photos from 29 different patients were selected. The lesions are from various locations on the body (face, neck, arm, trunk, leg, hand) and every patient has at least 2 images. From the 29 patients, 13 had photos taken from more than one appointment. In Figure 3.2 are shown examples of some of these photos, highlighting the diversity in camera position, lesion appearance and lighting.

Figure 3.2: Example of different photos taken by the doctor, showing the differences in acquisition setting and in hemangioma type and stage.

A binary ground-truth image was constructed for each image, differentiating the lesion area from the rest (Figure 3.3). For hemangiomas with multiple structures, the lesion area was considered as the region encompassing all structures. The final region of interest (ROI) images used for processing were obtained by considering the bounding box containing the lesion and enlarging it by 20%, as shown in Figure 3.4. This ensures that all ROIs dimensions are proportional to the lesion's dimensions and that the balance between non-hemangioma and hemangioma pixels is neither too high nor too low. If the number of non-lesion pixels were too high when compared to the number of lesion pixels, *Specificity* would be bias towards a good result; on the other hand, if the number of lesion pixels were too big when compared to the number of non-lesion pixels, *Sensitivity* would be bias towards a good result. Resizing the images to a standard size was considered, but due to the high variability in image size, resolution and lesion structures, it was not applied as it would severely distort most of the images (Figure 3.5). Also, one of the final goals in a CAD system of this kind is to calculate measures on the lesion or to assess its evolution on follow-up images, two aspects that would suffer from a standard image resize.

As stated above, a dataset of this kind presents advantages and disadvantages. The fact that there is no controlled acquisition setup means that the preprocessing step needs to be more thorough to balance the diversity in the images, and makes it more difficult for a good generalization of the segmentation methods. However, having various types of lesion stages, sizes, positions and resolutions, allows for the development of a more robust algorithm. Also, concerning the acquisition process, one should keep in mind that the patients are infants or small children, who are not always cooperative. This is even more critical if a rigid acquisition setup is necessary: they tend to move or even cry, specially if the lesion or the required position for acquisition causes pain or discomfort. Therefore, since consultations should be as swift as possible, the devised acquisition protocols should have these aspects into consideration.

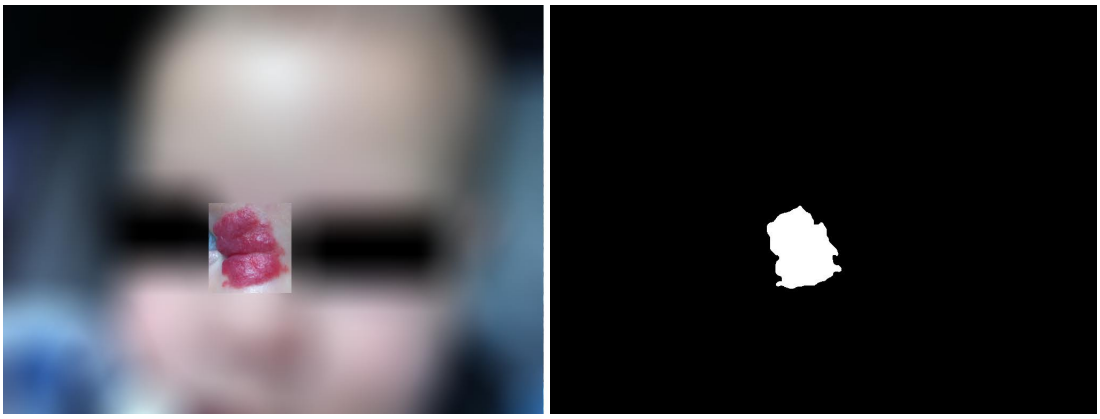

Figure 3.3: Original photo (left) and respective binary ground-truth image constructed (right).

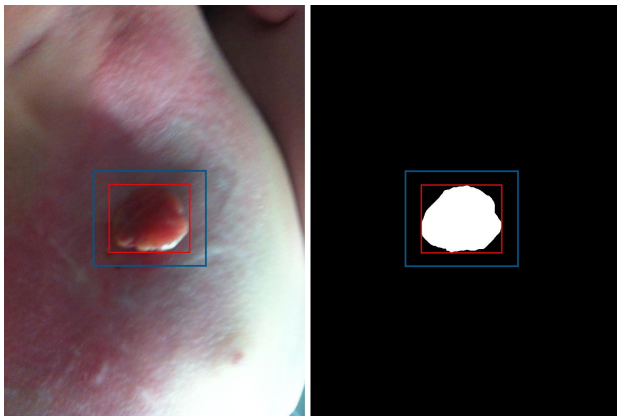

Figure 3.4: Bounding box surrounding the lesion (red) and the region considered for the new image, enlarging it by 20% (blue).

3.3 Preprocessing

Since images taken in unconstrained settings have large variability and may contain artifacts that hinder the segmentation process, the preprocessing stage is a fundamental step to achieve good results. To try and attenuate the heterogeneity of the images, an illumination and a color uniformization methods were employed, and a edge-preserving smoothing filter was applied to remove intrinsic noise and make the lesion region more homogeneous while still preserving its edges. For some steps and methods the images were converted to another color space that provides more useful channels than the standard RGB color space.

3.3.1 Color Space Transformation

Besides the RGB color space, the CIE 1976 L*a*b* (Robertson, 1977) space was also used. It has one channel for luminance (L^*) and two channels for color (a^* and b^*). It was designed to mimic the nonlinear response of the eye and therefore has perceptual uniformity, meaning that the difference between two colors can be viewed as the distance between them in a three-dimensional

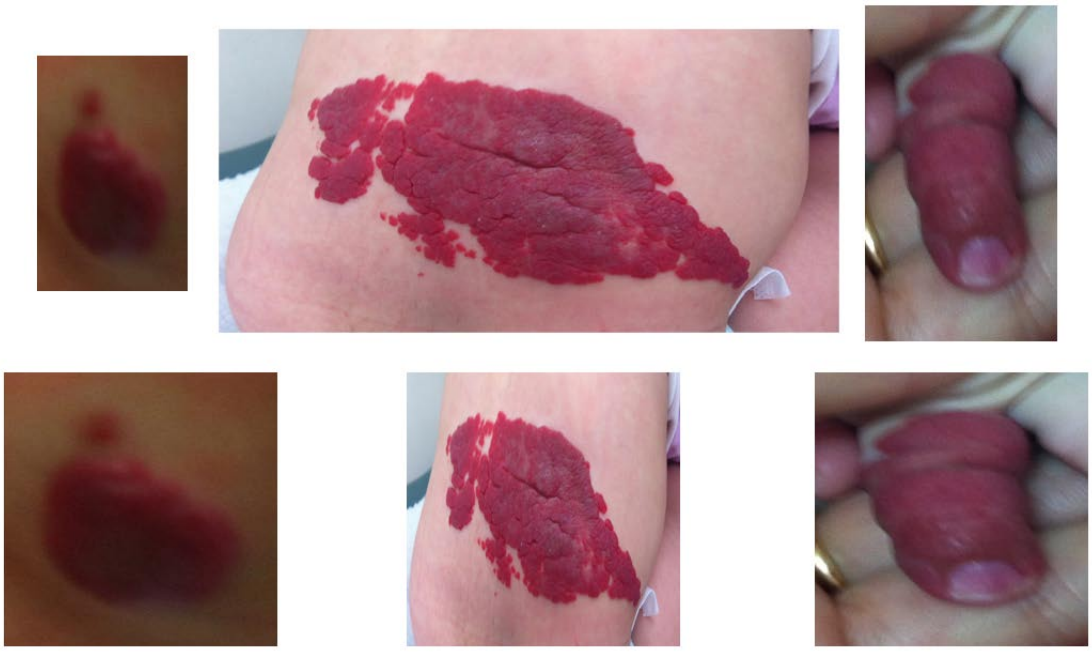

Figure 3.5: Three different examples of lesion structures affected by a standard rezise (256×256 pixels). Images on the first row are on their original size.

space. The L^* channel ranges from black at $L = 0$ to white at $L = 100$; the a^* and b^* channels are for color opponents green/red and blue/yellow, respectively. Positive values refer to red in the a^* channel and yellow in the b^* channel, while negative values refer to green in the a^* channel and to blue in the b^* channel. At $a = 0$ and $b = 0$, a neutral grey is achieved.

The L^* channel will be used to deal with uneven illumination, as discussed in the next subsection. Since the hemangioma lesions are red until they start to involute, a logic choice for processing the image would be a channel that highlights red colors, such as the a^* channel. In Figure 3.6 one can see the difference between the R channel from the RGB color space and the a^* channel. The separability of two distinct classes is more distinguishable on the a^* channel, as seen in the histograms. Therefore, this channel was chosen as input image for the majority of the segmentation methodologies. This was also a common preprocessing step presented in the state-of-the-art and achieved good results, which supports this decision.

As most of the segmentation techniques expect a single-channel input, other options like grayscale images, PCA representation of the image and other color channels were considered, but in the end, the a^* channel provided the best results. Oprisescu et al. (2015) also experimented with various color spaces in their work and conclude the same.

3.3.2 Illumination Uniformization

Due to the fact that the images were acquired with different lighting settings (flash/no flash, strong/weak natural light), this step aims to make all images in the dataset more similar concerning their luminance.

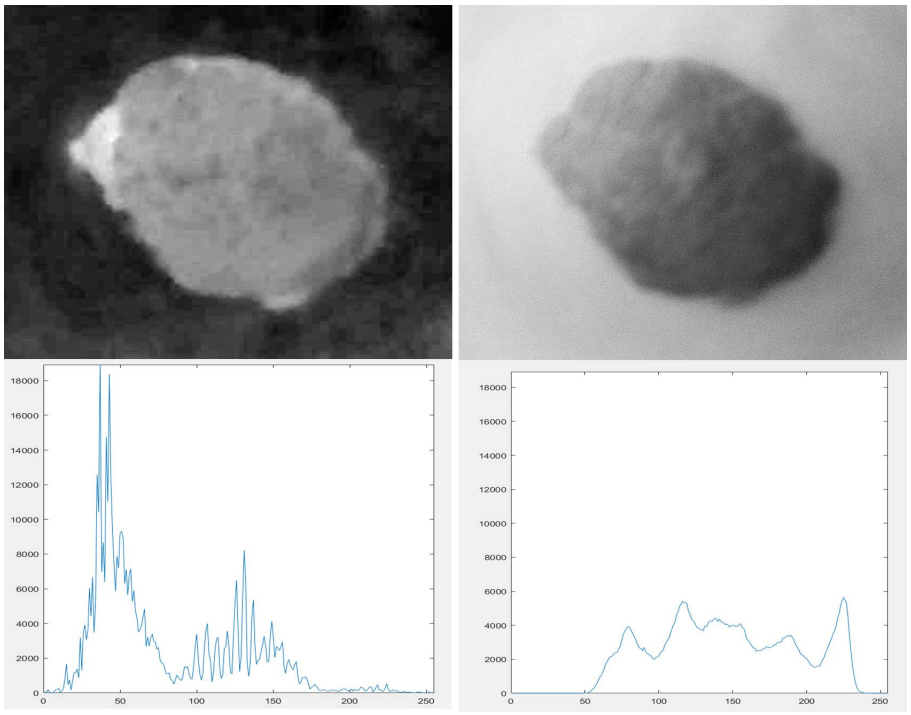

Figure 3.6: The a^* channel (left) and the R channel (right), taken from the same lesion. As one can see, the lesion and background are more distinguishable in the image on the left.

To achieve this, the input image was converted to the L*a*b* color space and the $L*$ channel retained. The histogram of this channel was computed, which represents the distribution of the image pixels in terms of every possible intensity value, organized in a set of 256 bins, one for each value. This was done for all images and the mean histogram calculated. To each image a transformation of the form $Y = T(X)$ is applied, to map the values in X to their new values in Y, so that the new histogram approximately matches the mean histogram. This transformation is chosen so it minimizes the following expression:

$$|c_{med}(T(k)) - c_i(k)| \ \ \forall k = 1,...,256 \tag{3.1}$$

where c_{med} is the cumulative sum of the mean histogram and c_i is the cumulative histogram of the image to be transformed. The new transformed luminance channel replaces the old one and the image is converted back to the RGB color space. In Figure 3.7 are represented the original images (far left), the histograms of the L^* channel (mid left), the histograms of the L^* channel after matching with the mean histogram (mid right) and the matched images back in the RGB space (far right). The three images have very different lighting conditions: the top-left image has weak lighting, the bottom-left image has strong lighting, and the middle-left image is somewhat in the middle. As one can see, after histogram matching, the image with low illumination has become lighter, the image with strong illumination has become darker and the middle image has not suffered many perceivable changes, as it did not have strong nor weak illumination.

Besides this method, the shading attenuation technique proposed by Cavalcanti et al. (2010)

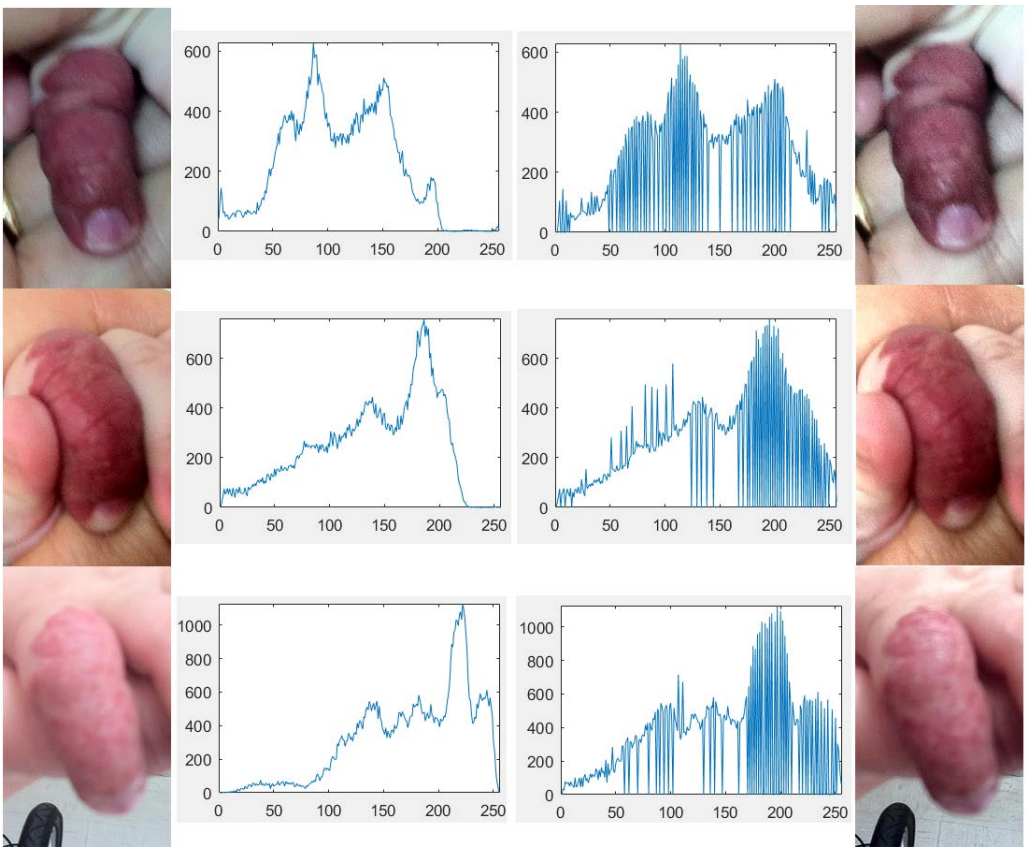

Figure 3.7: Original images (far left), histograms of the L^* channel (mid left), histograms after matching (mid right) and matched images (far right).

and the alpha blending method of Sultana et al. (2015b) were also tested, but the devised method yielded better results. The shading attenuation technique assumes that the lesion is centered and surrounded by an homogeneous background texture which has heterogeneous illumination. This is not the case for all smartphone photos of hemangiomas, as the lesion size and location may cause the inclusion of background regions that are not skin. As for the alpha blending approach, this method was devised to mitigate the effects of flashing on the images, by blending it with a less saturated mask. However, the variability in illumination may not be due to acquisition with the camera's flash on, but due to strong natural light.

3.3.3 Image Smoothing

To remove intrinsic noise that can be prejudicial in the segmentation phase, a smoothing filter is commonly applied to the image. The most widely used filters have an explicit kernel, such as the mean, median, Laplacian or Gaussian filters, for the purpose of blurring, shapening, extracting features or detecting edges. However, for segmentation tasks, preserving the edges is often an important step when considering smoothing operations. The Guided Filter (He et al., 2013) is a recent edge-preserving smoothing technique that computes the output filtered image by considering the content of a provided guidance image, which can be the input image itself or a different

one. It is defined by a local linear model between the output image q_i and a guidance image I_i, in a window w_k centered at a given pixel k such that:

$$q_i = a_k I_i + b_k \forall i \in w_k \tag{3.2}$$

where a_k and b_k are linear coefficients that are constant for w_k. These coefficients are determined such that the following cost function is minimized for the window w_k and the input image p_i:

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - p_i)^2 + \varepsilon a_k^2) \tag{3.3}$$

Equation (3.3) follows a Ridge regression model with ε penalizing large a_k and whose solution is given by:

$$a_k = \frac{\frac{1}{|w|}\sum_{i \in w_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \varepsilon} \tag{3.4}$$

$$b_k = \bar{p}_k - a_k \mu_k \tag{3.5}$$

A given pixel i will be considered by various overlapping windows w_k, and the output value of q_i will not be the same for different windows. Therefore, the average of all values of q_i is computed and Equation (3.2) becomes $q_i = \bar{a}_k I_i + \bar{b}_k$, where $\bar{a}_k$ and $\bar{b}_k$ are simply the average coefficients of the overlapping windows.

The result of applying this guided filter can be seen in Figure 3.8, where large areas are smoothed out of needless details and edges of the lesion are preserved.

Besides this method, another types of filters were tested: median, anisotropic diffusion and bilateral filter. However, those were discarded due to inferior performance and, in some cases, higher computational time when compared to the guided filter. Regarding the guidance image, a sharpened version of the input was considered to be used as guidance, but this would transfer noisy structures to the filtered image, and therefore was discarded.

Figure 3.8: Image before (left) and after guided filtering (right).

3.3.4 Color Constancy

The color of the hemangioma is an important characteristic that can be used in the segmentation step, as it displays a typical saturated red. The perception of color of an object depends on the light source and the surface properties, that have information of the reflected light. For this reason, an image of an object photographed under different light sources will show different colors. But humans have the ability to perceive the same colors under different illumination conditions, a feature called color constancy (Foster, 2011).

A number of color constancy algorithms have been proposed to attenuate the effects of varying illumination without having to have information of either the incident light or the surface reflectance (Barnard et al., 2002). One can define the image values $I = (R,G,B)^T$ over the visible spectrum ω, as dependent on the light source $e(\lambda)$, surface reflectance $s(\lambda)$ and camera sensitivity $c(\lambda)$, where λ is the wavelength, by:

$$I = \int_{\omega} e(\lambda)s(\lambda)c(\lambda)d\lambda \tag{3.6}$$

The goal is to estimate the projection of the light source color $e(\lambda)$ on the RGB space:

$$\mathbf{e} = \int_{\omega} e(\lambda)c(\lambda)d\lambda \tag{3.7}$$

Buchsbaum (1980) proposed the Gray-World hypothesis, that assumes that the average of a scene is gray, that is, the average reflectance is achromatic. Van De Weijer et al. (2007) propose a definition for this in the form:

$$\frac{\int s(\lambda,\mathbf{x})d\mathbf{x}}{\int d\mathbf{x}} = k \tag{3.8}$$

This constant k ranges from 0 (no reflectance) to 1 (total reflectance) over the domain of the entire image. With these assumptions, the reflected color is equal to the color of the light source, with:

$$\frac{\int \mathbf{I}(\mathbf{x})d\mathbf{x}}{\int d\mathbf{x}} = \int_{\omega} e(\lambda)c(\lambda)\frac{\int s(\lambda,\mathbf{x})d\mathbf{x}}{\int d\mathbf{x}}d\lambda = k\mathbf{e} \tag{3.9}$$

Another popular color constancy method called max-RGB (Land and McCann, 1971) assumes that a white patch exists in the image. Since a white patch would have maximum reflectance, it corresponds to the maximum of the RGB values:

$$\max_{\mathbf{x}} \mathbf{I}(\mathbf{x}) = k\mathbf{e} \tag{3.10}$$

$$\max_{\mathbf{x}} \mathbf{I}(\mathbf{x}) = \left(\max_{\mathbf{x}} R(\mathbf{x}), \max_{\mathbf{x}} G(\mathbf{x}), \max_{\mathbf{x}} B(\mathbf{x})\right) \tag{3.11}$$

The two methods above are just two instantiations of a more general approach based on the Minkowski norm. This p-norm is defined as:

$$\|X\|_p = \left(\sum_{i=1}^{N} |X_i|^p \right)^{1/p} \tag{3.12}$$

Finlayson and Trezzi (2004) called this more general method Shades of Gray, which is given by:

$$\left(\frac{\int (\mathbf{I}(\mathbf{x}))^p \, d\mathbf{x}}{\int d\mathbf{x}} \right)^{1/p} = k\mathbf{e} \tag{3.13}$$

As one can see from Equation (3.13), if $p = 1$ it becomes the Gray-World algorithm, and if $p = \infty$ then it becomes the max-RGB algorithm. The authors found that the best results are obtained with a Minkowski norm of $p = 6$.

To achieve the desired color constancy across all images, the Shades of Gray algorithm with $p = 6$ was applied to all images, with some results presented in Figure 3.9. The reason for this choice is that the algorithm is based on low-level features of the image, easy to implement, fast and with good results.

Figure 3.9: Four examples of different hemangiomas that appear to have been acquired in similar conditions.

3.4 Methodologies for Lesion Segmentation

In this section, the various segmentation methodologies used in this work are explained, as well as the various steps before and after segmentation.

Here, we divide the approaches in shallow and deep methodologies, that is, those that base the segmentation on low-level, fixed features of the image, and those that learn how to segment the image based on low and high level features, reaching more abstract representations of the image.

3.4.1 Shallow Methodologies

In this work, shallow methodologies will be referring to segmentation approaches that work with low-level, fixed image features, such as image statistics, pixel intensity or image gradients. Various types of methods were used, which can be divided into thresholding-based methods, edge-based methods, active contour-based methods and artificial intelligence-based methods.

As stated in section 3.2.1 on color space transformation, the reddish color of the hemangiomas is highlighted in the a^* channel of the L*a*b* color space. Therefore, all preprocessed images are converted to the L*a*b* color space and the a^* channel is converted to the range $[0,1]$ and used as input image I for all the methodologies used in this subsection.

3.4.1.1 Thresholding using Otsu's Method

Thresholding techniques make use of one or multiple thresholding values that separate the regions of interest (ROI) in the images. The simplest form makes use of a value t to binarize the image such that if $I_{i,j} \geq t$ then $I'_{i,j} = 1$, else $I'_{i,j} = 0$ for any given image pixel (i, j). It is a commonly used method, as usually the ROI and the background are heterogeneous in such a way that can be easily separated by just one value of t.

One of the most widely used methods to determine the value of t is Otsu's method (Otsu, 1979). It searches for the optimal threshold that divides the normalized histogram of k bins into two homogeneous classes, $C_0 = [1,2,...,t-1]$ and $C_1 = [t,t+1,...,k]$, by minimizing the intra-class variance σ_b^2, defined by:

$$\sigma_b^2 = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \tag{3.14}$$

$$\omega_0 = \sum_{i=1}^{t-1} P(i),\ \omega_1 = \sum_{i=t+1}^{k} P(i) \tag{3.15}$$

$$\mu_0 = \sum_{i=1}^{t-1} i\frac{P(i)}{\omega_0},\ \mu_1 = \sum_{i=t+1}^{k} i\frac{P(i)}{\omega_1},\ \mu_T = \sum_{i=1}^{k} iP(i) \tag{3.16}$$

In Equations (3.15) are defined the class probabilities, ω_0 and ω_1, and in Equations (3.16) the class means, μ_0 and μ_1 and the input image total mean intensity, μ_T.

Despite being an easy and fast technique to implement, the segmented images tend to be over- or under- segmented due to poor contrast between the lesion and the background, or with very irregular edges. For this reason, a series of morphological operations often need to be performed, such as hole filling or elimination of objects with small area that were erroneously segmented. All these operations will be discussed in section 3.5.

3.4.1.2 Edge Detection with Canny's Algorithm

One common approach in image processing for the segmentation of skin lesions is to use edge detectors, that can separate the lesion's borders from the background. One of such methods is

the Canny edge detector (Canny, 1986), widely used due to its advantages over other edge detectors: avoids detecting double edges, the detection has low error probability and allows for a good location of the edge pixels.

The algorithm consists of 5 stages. First, the image is smoothed with a Gaussian filter $G(x,y)$ to prevent false detections, since the results are easily affected by image noise. The input image $I(x,y)$ is convolved with this filter:

$$I^{'}(x,y) = I(x,y) * G(x,y) \tag{3.17}$$

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}} \tag{3.18}$$

where σ is the Gaussian function standard deviation. The next step is to find the gradient magnitude $M(x,y) = \sqrt{M_x^2 + M_y^2}$ and direction $\theta = \arctan\frac{M_y}{M_x}$ of each pixel in the smoothed image. Since the edges extracted at this stage are still quite blurred, a non-maximum suppression technique is applied to retain only the local maxima, that is, the location of the pixels with the sharpest change in intensity values. This is done by comparing the edge strength of a given pixel with the edge strength of the pixels in the negative and positive gradient directions; if it is largest when compared to the others, the value is preserved, otherwise is set to 0.

Due to color variations and noise, there are still some false edge pixels detected at this stage. Therefore, the next step consists of a double threshold (T_1, T_2) to identify strong and weak edges. If a given edge pixel has a value above T_1 it is marked as a strong edge; if its value lies between T_1 and T_2, it is marked as a weak edge; if its value is smaller than T_2 it is suppressed. The values for these two thresholds depend on the content of the image.

The final stage consists on determining if the weak edges are indeed part of the image's final edges. This is done by analyzing the 8-connected neighborhood pixels of these edges: if the weak edge pixel is connected to a strong edge pixel, then it is marked to be preserved.

This method presents two major problems: the detected edges are usually disconnected, which means that the border of the lesion is not fully detected and there is a large sensitivity to noise that results in false edges. A postprocessing step to connect the detected edges and fill the boundaries is usually applied, when a closed object is desired. This will be discussed in section 3.5.

3.4.1.3 Fuzzy *c*-Means Clustering

Clustering refers to the subdivision of a given dataset S into c subsets that are disjoint, non-empty and that constitute S by union. In fuzzy clustering (Bezdek et al., 1984), the data points can belong to more than one cluster. The method is based on the minimization of the following objective function:

$$J_m = \sum_{i=1}^{D}\sum_{j=1}^{N} \mu_{ij}^m \|x_i - c_j\|^2, \;\; m \in \mathbb{R} \;\; \text{with} \;\; m \geq 1 \tag{3.19}$$

where D is the number of data points in S, N is the number of clusters, μ_{ij} is the membership degree of x_i in cluster j, m is the level of fuzziness of the boundaries between clusters, x_i is the ith data point of S and c_j is the center of the jth cluster.

The clustering algorithm starts by randomly initialize the membership values μ_{ij} and then calculates the cluster centers using:

$$c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^m x_i}{\sum_{i=1}^{D} \mu_{ij}^m} \tag{3.20}$$

The new centers are used to update the membership function:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{N} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{3.21}$$

The value of the objective function J_m is calculated using the centers and the updated membership values. This is repeated until the value is smaller than a given threshold, $J_m < \varepsilon$, or after a specified number of iterations. After convergence of Equation (3.19) to a local minimum, the membership value of each data point to each cluster is inspected, and that point is attributed to the cluster with the highest value.

This algorithm can be used to segment skin lesions, by selecting $c = 2$, one cluster for the lesion and other for the background. However, for images with low contrast it may happen that the lesion pixels are not clustered as belonging to the lesion, or that some skin pixels may be clustered as lesion. Again, a postprocessing step is vital to ensure that only one object is retained.

3.4.1.4 Boundary Detection with Gradient Vector Flow Snakes

Methods based on active contours have been used in the segmentation of skin lesions by various authors (Silveira et al., 2009; Zhou et al., 2010). The principle of active contours relies on a initial curve that is iteratively deformed to adapt to the boundaries of an object of interest. The basic snake model (Kass et al., 1988) consists of a spline that is under the influence of an internal elastic energy and an external edge-based energy, and can be defined parametrically as $\mathbf{v}(s) = [x(s), y(s)]$, where $s \in [0,1]$ minimizes the energy function:

$$E = \int_0^1 E_{int}(\mathbf{v}(s)) + E_{im}(\mathbf{v}(s)) + E_{con}(\mathbf{v}(s))ds \tag{3.22}$$

In this equation, E_{int} represents the bending energy to which the snake is subject to, E_{im} the energy due to the intensity of the image and E_{con} the constraint energy that can be user-defined or automatically established.

The Gradient Vector Flow (GVF) snake (Xu and Prince, 1998) uses a GVF field as the E_{con} in Equation (3.22). To get this GVF field, the first step is to compute the edge map of the image, that can be obtained with the function:

$$f(x,y) = -|\nabla I(x,y)|^2 \tag{3.23}$$

where ∇ is the gradient. We can then compute the GVF field $\mathbf{g}(x,y) = (u(x,y), v(x,y))$ that is defined as the equilibrium solution that minimizes the following energy function:

$$\varepsilon = \iint \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |\mathbf{g} - \nabla f|^2 dxdy \tag{3.24}$$

After obtaining $\mathbf{g}(x,y)$ and replacing in Equation (3.22) as the E_{con} term, the snake can be iteratively deformed until E converges. The initialization of the snake can be done by the user or in an automatic manner, and can be far from the boundaries or across them.

One disadvantage of this method relies in the edge map and subsequent GVF field having weak intensities and thus not being able to push the snake to the boundaries of the object, which result in final contours that are very similar to the snake initialization. However, since the snake initialization is a closed contour, one can always expect a closed contour. The postprocessing step can fill the contour so that the final result is a complete closed object.

3.4.1.5 Segmentation with Closed Shortest Path

The problem of detecting edges can be viewed as finding the minimum cost path in a directed and weighted graph that represents a closed contour. A graph G is composed of a set of vertices or nodes, V, and a set of arcs (p,q), A, such that $G = (V,A)$. The graph is said to be directed if the set A is a set of ordered pairs of nodes, and to be weighted if a weight $w(p,q)$ is associated with each arc. For images, $w(p,q)$ is usually a function of the pixels values and their relative positions. A path in the image is a set of unique vertices $v_1, v_2, \ldots, v_n$, where v_i and v_{i+1} correspond to neighbor pixels, and the total cost is defined as the sum of each of the arcs in the path, $C = \sum_{i=2}^{n} w(v_{i-1}, v_i)$. The cost of each edge is defined in this approach as a function of the features computed at the head of the arc and the distance to the causal neighbors (between the head and the tail of the edge). A causal neighbor is defined as the subset of neighbors v of a node p, such that v_i precedes p (Figure 3.10).

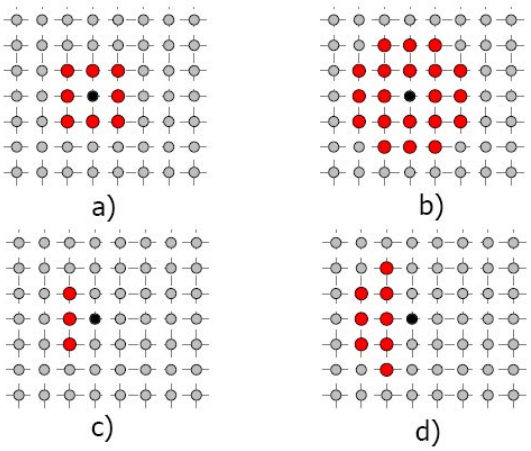

Figure 3.10: Example of a given node (black) with its 8 neighbors (a) and 20 neighbors (b). In (c) and (d) are represented the causal neighbors considering 8 (c) and 20 (d) neighbors, respectively, in the left to right direction. Adapted from (Cardoso et al., 2015).

Given a point that is enclosed by a boundary, one can transform the image to polar coordinates using that point as origin. By doing this, the closed shortest path results in the shortest path between the opposite margins, given that the start and end points are constrained to be neighbors. In this form, the ordering of vertices is given by the column number and the causal neighbors of a node are those with a lower column number.

Cardoso et al. (2015) propose a novel contribution to this technique by finding the minimum cost path enclosing a seed point O without having to transform the image to polar coordinates. The nodes need to be ordered by angle θ relatively to point O, such that $\theta \in [0, 360]$. To define the edges of the path to be obtained, one needs the causal neighbors of a given node. For a given node p positioned at an angle θ relative to the origin point O, the causal neighbors of p are the nodes v_i in the d-neighborhood of p with $\theta_i < \theta$. This implies that the number of causal neighbors changes from node to node, depending not only on θ but on the position of that node in the ROI.

An important point when considering circular paths in the original coordinates is that enclosing paths closer to point O are shorter (by number of edges), than those that are further from the seed point. If the edge cost is simply based on features extracted from the edge or in the pixel neighborhood, then smaller contours will be preferred since they have less edges and therefore smaller cost. As this is not something to favor, the authors weight the cost of an edge by $1/r$, r being the distance from the seed point to the head node of the edge.

As stated before, the cost of an edge is modulated by a factor related to features computed at the head of the edge and a factor concerning the distance between the head and tail. The features computed at the edge regard the gradient along a radial direction, θ, calculated as a 3-point differentiation:

$$G_\theta = \frac{f(r+h) - f(r-h)}{2h} \tag{3.25}$$

where $f(.)$ is the intensity value along the radial direction, r is the radius and h was set to 1. The weight of the edge is a nonlinear function of the derivative, expressed as an exponential law:

$$w(g) = f_l + (f_h - f_l)\frac{e^{\beta(255-g)} - 1}{e^{255\beta} - 1} \tag{3.26}$$

where g is the magnitude of the derivative, $f_l = 2$, $f_h = 32$ and $\beta = 0.025$, as set by the authors.

Concerning the distance factor between the head and the tail of the edge, the authors found that using a polar distance metric would produce the best results when compared with other metrics. Therefore, d_{pol} is defined using a resolution of one degree and one radial unit per pixel as:

$$d_{pol} = \sqrt{(\Delta r)^2 + (\Delta\theta)^2} \tag{3.27}$$

One disadvantage of this method is the need for a seed point, that can be user-defined, or automatically calculated as the center of the image (if the object to segment is centered) or the object center of mass. The method always computes a closed contour, so no disconnected edges are to be expected, and the boundary can be filled to result in a complete filled object.

3.4.2 Deep Learning Methodologies

Another approach evaluated for lesion segmentation was Deep Learning. Deep Learning has emerged in the last decade as a new area in Machine Learning that has since made major advances in solving many computer vision and artificial intelligence problems that previous methods could not solve (Krizhevsky et al., 2012; LeCun et al., 2015). It thrives due to being able to learn multiple levels of data representation that correspond to multiple levels of abstraction, where high-level concepts are defined from low-level ones. The key aspect for this are the models that consist of multiple layers or stages of nonlinear processing with methods that can learn the higher level feature representations. In this work, since ground truth images are available, learning was made in a supervised manner using deep networks, specifically Convolutional Neural Networks (CNN), whose theoretical background will be thoroughly explained in the next subsections.

Before tackling the more complex structure of CNNs, it is best to explain first standard Neural Networks. A neural network is composed of layers of neurons that connect to other neurons in the next layer, forming a organized structure. To each neuron is associated a set of N inputs x_i and weights w_{ij}, and a bias b_j, which are linearly combined in the form:

$$a_j = \sum_{i=1}^{N} w_{ij} \cdot x_i + b_j \tag{3.28}$$

where a_j is called the activation of the jth neuron. Each activation is then transformed using a nonlinear activation function $h(.)$ that produces the output of a given neuron $z_i = h(a_j)$. This activation function can be the hyperbolic tangent function, the logistic sigmoid function or the more recent ReLU units. These various outputs z_i are again linearly combined and transformed, ultimately giving rise to the network outputs y_j.

It is this kind of architecture that composes CNNs, but with different types of layers to process particular features of the image. The simplest architecture is comprised of pairs of convolutional layers and pooling layers, as shown in Figure 3.11.

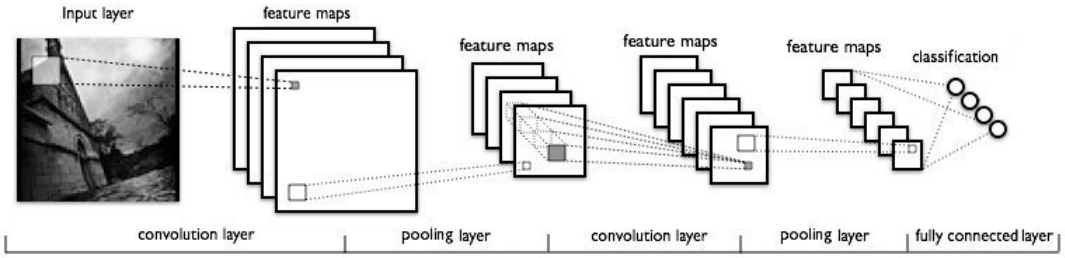

Figure 3.11: A simple CNN architecture.

3.4.2.1 Convolutional Layer

This is the foundational layer of a CNN. It differs from a standard neural network layer in the fact that the inputs are not all connected to every neuron in the next layer. This would lead to thousands

of weights and bias that would increase exponentially when working with high resolution images, leading to unbearable computational complexity. Rather, this type of layer consists of a number of filters K that have a spatial area F with a weight and bias associated. These filters slide horizontally across the input, convolving with the image and computing the dot product. The result is a 2D activation map for that particular filter, as shown in Figure 3.12. Taking the example of the first convolutional layer in a CNN, the input is an image with dimensions $W \times H \times D$ and every filter will perform this operation along dimension D, which means that the depth of the filters must be equal to the depth of the input data. The activation map of each filter is stacked depth-wise, producing the output of the convolutional layer of size $W \times H \times D$, where $D = K$. These filters will activate for specific features in the image, meaning that each filter only detects a single kind of localized feature.

Two other important factors in this layers are the amount of stride and zero-padding. The stride S is the number of steps that the filter will slide before every convolution. Zero-padding is employed as a way of solving the problem of the size of the filter not being able to completely cover the input data, by adding P pixels with value 0 along the border of the input.

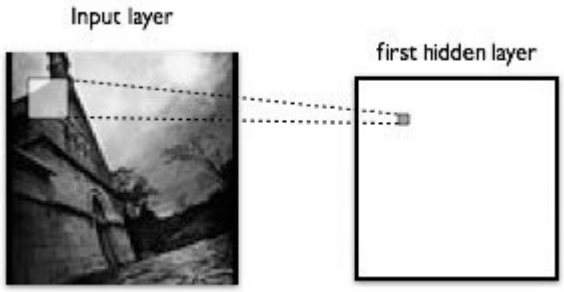

Figure 3.12: Convolutional layer with a region of an image connected to a neuron.

3.4.2.2 Rectified Linear Unit

The Rectified Linear Unit (ReLU) implements a rectifier, an activation function defined as $f(x) = \max(0,x)$. This unit basically takes the output of a neuron x and sets the value to 0 if x is negative or leaves it as it is if otherwise. Russakovsky et al. (2015) showed that this type of units present a training process several times faster than with the tanh function.

3.4.2.3 Pooling Layer

Pooling layers perform downsampling to the input received to reduce the size along the first and second dimensions. The most common type of pooling used is max pooling. In this type of layer, a $p \times p$ region of the feature map generated by the convolutional layer is considered and pooled, outputting the maximum value in that region, as shown in Figure 3.13. This generates a more condensed version of the map, with the advantage of decreasing the amount of storage and processing needed in subsequent layers.

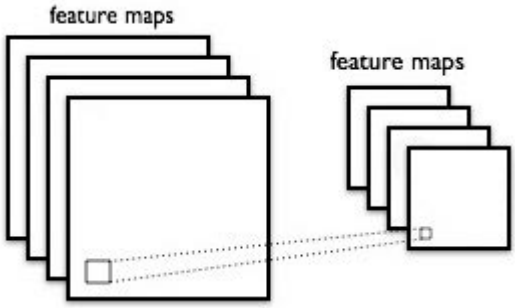

Figure 3.13: Pooling layer outputting a condensed feature map.

3.4.2.4 Fully Connected Layer

This is a special case of the convolutional layer, where all the output activations of the previous layer are connected to all the neurons in this layer (Figure 3.14). The goal is to produce the high level abstraction of information. Due to these neuron connections, the spatial relation between the feature maps and the image is lost, meaning that no more convolution layers can be added after this layer.

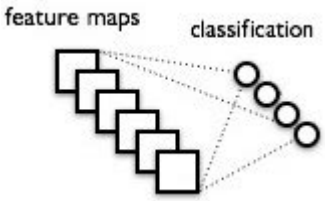

Figure 3.14: Fully connected layer, where all the computed feature maps are connected to the output neurons.

3.4.2.5 Loss Layer

The loss layer is, similarly to the ReLU units, the activation function of the fully connected neurons that computes the output of the network. It determines how the network training will penalize the deviation between the true and predicted labels. One of the most common used loss functions for CNNs is the softmax loss. It combines the softmax function with the logarithmic loss function:

$$l(x,c) = -\log \frac{e^{x_c}}{\sum_{k=1}^{C} e^{x_k}} \tag{3.29}$$

where x is the output vector of probabilities from the network, c is the class for which the loss is calculated and C is the number of classes. This type of function is preferred for its probabilistic interpretation of the output.

3.4.2.6 Training

To make a network able to correctly classify an input, the filter's weights and bias have to be adjusted. This is done by forward and backward propagation. First, an image runs through the network producing as output a score vector that contains the probabilities of a given class for that image - forward propagation. This score is compared with the true value to compute the error using the defined loss function, that is, the deviation between the true and predicted scores. Since we want the error as small as possible, this turns into a minimization problem. Backward propagation is used to compute the error's derivatives with regards to the weights by applying the chain rule. These are used to update the weights in the direction of the steepest descent to make the error as small as possible. One common used method to achieve this is the Stochastic Gradient Descent (SGD) method.

Using a simple example, consider a layer $y = f(x,w)$, where x is the input and w the associated weight, that after a forward propagation produces a value y with an associated loss l. The gradient of l with respect to y is known and given by $\frac{\partial l}{\partial y}$. We want to know how the loss relates to the parameters of y, $\frac{\partial l}{\partial x}$ and $\frac{\partial l}{\partial w}$. This can be achieved through the chain rule: $\frac{\partial l}{\partial w} = \frac{\partial l}{\partial y}\frac{\partial y}{\partial w}$. The weights can then be updated to be used in the next iteration with the following equation:

$$w^n = w^{n-1} - \eta \frac{\partial l}{\partial w} \tag{3.30}$$

where η denotes the learning rate and n is the number of the epoch.

In summary, to train the network, a batch of the data is fed to the network to produce a classification and loss. Using the loss, a back propagation is made to calculate the weight gradients and the network parameters are then updated by SGD. This is repeated to, hopefully, make the CNN converge to a small loss error, in a number of iterations called epochs. One part of the data is used for training and the other is used for validation. As the name implies, only the data in the training set affects the weights and the validation set is used to assess if the training is making the network converge and to detect any problem or failure.

3.4.3 Transfer Learning from Melanoma

As discussed in Chapter 2, the development of methodologies specific for the segmentation of IHs is a relatively unexplored topic. However, from a domain point of view, this falls into the category of segmenting skin lesions, which has a wider range of tasks and problems to be solved, but that is still circumscribed to a specific type of data (concerning imaging techniques, color gamut or expected features). The question of whether or not, solutions devised for solving a specific problem regarding a type of skin lesion can be adapted and applied to another type, raises itself as an important and interesting point. Most of the machine learning problems have good results under the supposition that the data used for training and testing have the same feature space and similar statistical distributions. What if one changes the acquisition process/sensor for the same task? What if the task changes slightly, but the core problem remains similar? These models would have to be rebuilt using new data, something that in many real world applications would

be impossible or very expensive to do. This is even more pronounced for Medical Imaging, as the datasets are very scarce and difficult to construct. This is where *Transfer Learning* comes into play.

Transfer Learning can be defined as an approach where we solve a new task using knowledge acquired in another, without having to redo all the training procedure. Following the definition of Pan and Yang (2010): given a source domain D_S and a learning task T_S, a target domain D_T and a learning task T_T, transfer learning aims to help improve the learning of the target predictive function $f_T(.)$ in D_T, using the knowledge in D_S and T_S, where $D_S \neq D_T$ or $T_S \neq T_T$. In Figure 3.15 is depicted the comparison between how models are trained in traditional machine learning and with transfer learning. In the first, we usually have datasets (represented by the circles) with enough data to generate models; in the other, the data is used to generate a model aimed at solving a task and then the knowledge gained in training that model is transferred to generate a model for solving another task for which we do not have enough data.

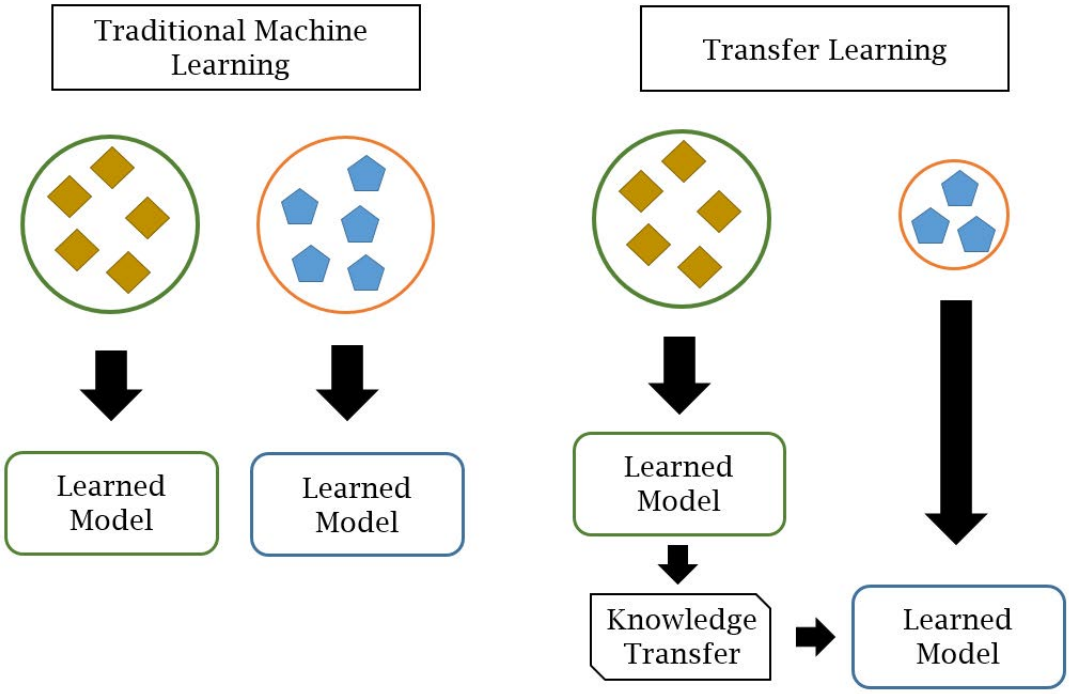

Figure 3.15: Comparison between the classic machine learning training procedure and the transfer learning procedure.

Only recently has transfer learning been used to solve tasks related to skin lesions. Fornaciali et al. (2016), Lopez et al. (2017) and Menegola et al. (2017) used pre-trained, publicly available CNNs and fine-tuned them using new datasets, for the purpose of melanoma screening; Liao (2016) goes further and tries to develop a universal skin disease diagnosis system, fine-tuning pre-existing CNNs with images from various types of skin lesions; for the purpose of segmentation, Qi et al. (2017) address the problem of melanoma segmentation by fine-tuning a pre-existing CNN and then transforming it to a Fully Connected Network.

In this work, transfer learning will be done in a supervised way. The goal is to train a CNN on a set of melanoma images and then make use of that knowledge by fine-tuning the CNN to be applied to hemangioma lesions. This means that, instead of randomly initialize the weights and bias of all the convolutional layers when training the CNN with IH images, those initializations come from the training done with melanoma images. In summary, in this supervised transfer learning approach, we transfer source weights to fine-tune the target network.

3.5 Postprocessing

Due to the lack of contrast between the lesion and the background or the presence of noise that the preprocessing step was not able to mitigate, the result segmented images can often be oversegmented or undersegmented.

Threshold and clustering techniques often result in oversegmented images, having many small regions erroneously segmented. One way of counter this is to apply some morphological operations to end up with only the segmented lesion.

Morphological operations use a binary or grayscale image as input along with a structuring element and combine them using set operations (intersection, union, complement, inclusion). The objects in the input image are processed based on the shape that is related to the structuring element. Since all morphological operations in this work were applied to binary images, the following definitions and examples are all related to this type of images.

The basic operations in mathematical morphology are *erosion* and *dilation*. Erosion, as the name implies, refers to an operation that erodes objects that belong to the foreground (i.e. white pixels), making the area of these objects shrink. Dilation does the opposite, as it enlarges objects in the foreground, making the area of these objects increase. From this two operations one can derive the operations of *closing* and *opening*. Opening is defined as an erosion followed by a dilation using the same structuring element for both operations. This operation can counter the undesired effect of applying an erosion, which affects all structures indiscriminately and not only "salt noise". By applying a dilation after, it ensures that all pixels that fit in the structuring element that belong to the foreground are preserved. Closing is the opposite, defined as an dilation followed by an erosion using the same structuring element for both, and serves the opposite goal than that of the opening operation.

These operations are useful to refine the segmentation results. When detecting edges, is sometimes difficult to obtain a closed contour. The dilation and closing operations are useful to connect disjointed edges. When applying threshold and pixel-wise classification methods is common to obtain objects that do not belong to the object to be segmented. In this case, the erosion and opening operations are useful to eliminate most of these without affecting the main object.

3.6 Performance Metrics

To evaluate the segmentation results obtained with the various methods, performance metrics are used, which compare the results to the ground-truth image.

As the classification of the pixels in the image is binary, one can define the concepts of true/false positives and true/false negatives, which make up a confusion matrix as shown in Table 3.1.

Table 3.1: Confusion matrix defining true/false positive/negative classified pixels.

		Segmented Pixels	
		Lesion	Background
Ground Truth Pixels	Lesion	True Positives (TP)	False Negatives (FN)
	Background	False Positives (FP)	True Negatives (TN)

From these definitions is possible to establish some performance metrics. The ones used in this work were: Accuracy, Sensitivity, Specificity, Dice Coefficient, Jaccard Index, XOR measure and Haussdorf Distance. The purpose of such wide array of metrics was to establish some state-of-the-art reported results that did not existed before concerning IH segmentation, but also to produce results with metrics that are commonly used. Therefore, in the next subsections, the definitions of each metric is provided.

3.6.1 Accuracy

Accuracy measures the fraction of pixels that are correctly segmented, and is given by:

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \tag{3.31}$$

3.6.2 Sensitivity

Sensivity or True Positive Rate, measures the proportion of lesion pixels that are correctly identified as such, given by:

$$SE = \frac{TP}{TP+FN} \tag{3.32}$$

3.6.3 Specificity

Specificity or True Negative Rate, measures the proportion of background pixels that are correctly identified as such, given by:

$$SP = \frac{TN}{FP+TN} \tag{3.33}$$

3.6.4 Dice Coefficient

The *Dice coefficient* measures the similarity of two segmentations, giving more weight to instances where the segmentations agree. It can be expressed by:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{3.34}$$

3.6.5 Jaccard Index

The *Jaccard index* is a measure of similarity between two segmentations that computes the ratio between the intersection and the union of the segmented image, $I_S and ground - truth image, I_{gt}$:

$$J_{ind} = \frac{|I_{gt} \cap I_S|}{|I_{gt} \cup I_S|} \tag{3.35}$$

But can also be expressed in relation to the concepts derived from the confusion matrix. It relates the TP to the number of pairs that are in the same cluster or belong to the same class, neglecting the TN:

$$J_{ind} = \frac{TP}{TP + FP + FN} \tag{3.36}$$

3.6.6 XOR measure

The *exclusive disjunction (XOR) measure* is the level of disagreement between the segmented image and the ground truth image:

$$XOR = \frac{|I_{gt} \oplus I_s|}{|I_{gt}|} \tag{3.37}$$

where $\oplus$ denotes the exclusive disjunction (XOR) operation. This metric can also be expressed in relation to the concepts derived from the confusion matrix:

$$XOR = \frac{FP + FN}{TP + FN} \tag{3.38}$$

3.6.7 Haussdorf Distance

The *Hausdorff distance* measures the extent to which each point of a set lies near some point of another set. Therefore, it can be used to measure the degree of resemblance between segmented borders and the respective ground-truth.

Given the set of points that constitute the segmented border, $S = s_1, s_2, ..., s_n$, and the set of points that constitute the ground-truth border, $G = g_1, g_2, ..., g_m$, the minimum distance between each point of S and any point of G is computed, and the maximum obtained value is retained,

$h(S,G)$; the reciprocal is also computed, with the maximum value retained, $h(G,S)$. The Hausdorff distance corresponds to the maximum value between $h(S,G)$ and $h(G,S)$.

$$HD(S,G) = \max(h(S,G), h(G,S)) \tag{3.39}$$

$$h(S,G) = \max_{s \in S} \min_{g \in G} \|s - g\| \tag{3.40}$$

Chapter 4

Results and Discussion

In this chapter, we present the results obtained with the methods described in the previous chapters. Each section will detail the workflow of each approach as well as the parameters used. For each method the results are discussed, highlighting strong and weak points. Lastly, a wrap-up of all the experimented frameworks is summarized.

4.1 Thresholding using Otsu's Method

To segment the lesions by thresholding, first they are preprocessed using the methods detailed in section 3.3. Then, as already discussed in section 3.3.1, and derived from the reviewed state-of-the-art approaches, the input images are converted to the L*a*b* color space and the a^* channel is retained and used as input image. A threshold t is computed for each image using Otsu's method, which is then used to binarize the image. To refine the segmentation, some morphological operations and hole filling are applied and the final result is obtained by retaining only the largest object. The complete workflow for this approach can be seen in Figure 4.1. A few segmentation results are showed in Figure 4.2, for different types of lesions. As one can see, in some images the threshold oversegments the lesion (Figure 4.2a and 4.2c), derived from the low contrast between the border of the hemangioma and the surrounding skin; in images of lesions that are undergoing involution, the method results in undersegmentation (Figure 4.2b), labeling the pixels of these involuting areas as non-hemangioma.

Therefore, postprocessing methods are applied to counter these adversities. The operations of opening and erosion were used with a structuring element that depends on the size of the input image, followed by hole filling. The final image is obtained by considering only the biggest object remaining. In Figure 4.3 are shown the results of this processing, applied for the same images. An increase in performance was observed, both visually as well as quantitatively. In Table 4.1 are presented the mean results obtained for each metric for this method, after applying postprocessing.

However, as expected, this is not a flawless method, nor the postprocessing can make it perfect. In Figure 4.4 we depict some of the cases with poor performance. As referred above, the

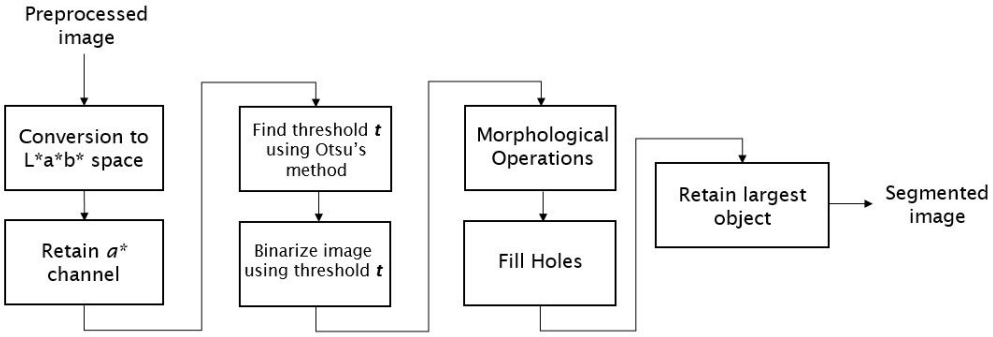

Figure 4.1: Block diagram of the framework devised for using Otsu's method.

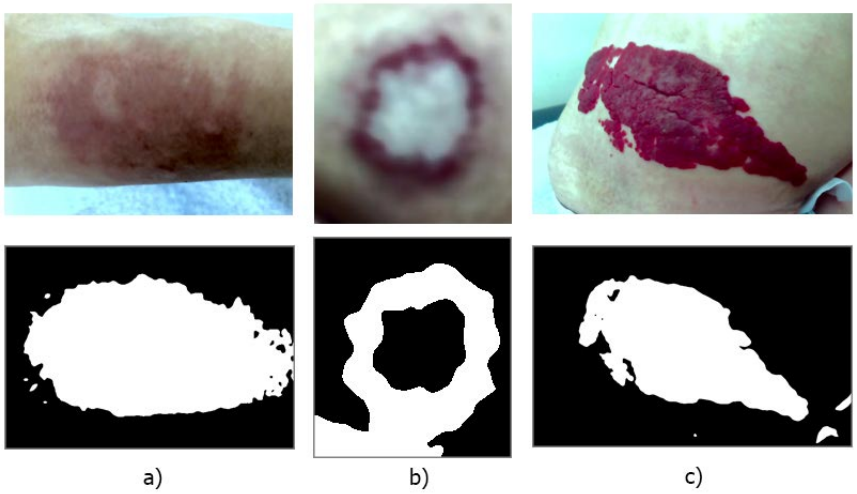

Figure 4.2: Three examples of segmentation using Otsu's method.

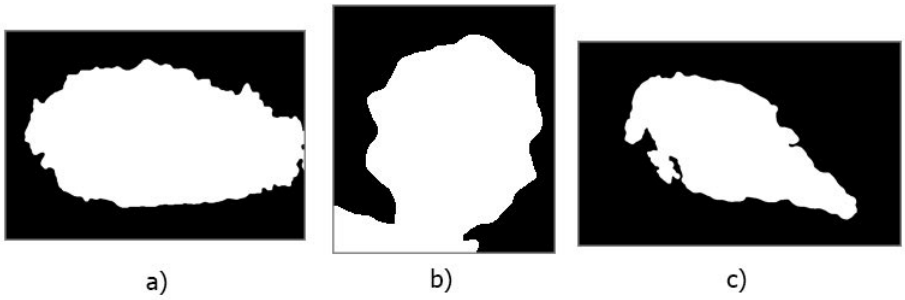

Figure 4.3: Results obtained after applying postprocessing.

Table 4.1: Mean results obtained with Otsu's method.

ACC (%)	SP (%)	SE (%)	HD (px)	XOR (%)	Dice (%)	Jacc (%)
84.30	89.58	81.57	138.92	45.26	81.05	70.40

method has difficulties when the lesion border has low contrast with the skin, or when the lesion is involuting.

Figure 4.4: Examples showing poor segmentation results with Otsu's method.

4.2 Edge Detection with Canny's Algorithm

The detection of borders using Canny's algorithm was made using the a^* channel of the image after conversion to the L*a*b* color space, as seen in the block diagram of Figure 4.5. The standard deviation of the Gaussian function used to filter the image was set empirically for each image and the values of the double thresholds T_1 and T_2 were also set empirically for each image. In Figure 4.6 one can see some examples obtained with this method.

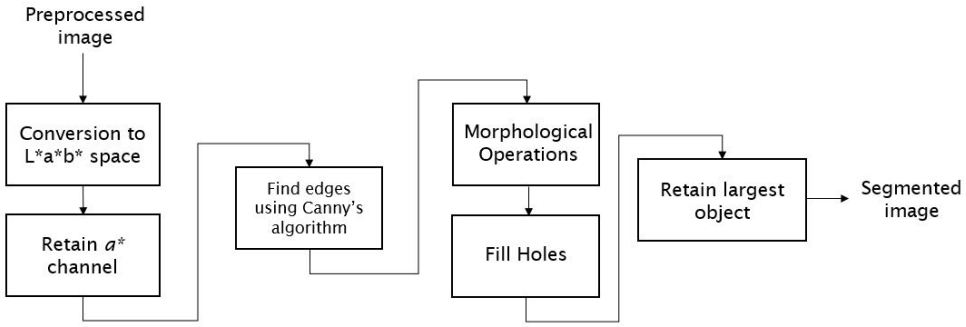

Figure 4.5: Block diagram of the framework devised for using Canny's Algorithm.

As shown, it is difficult to obtain a closed contour simply by applying the method (Figure 4.6). Therefore, to obtain a well defined border and complete lesion object, the postprocessing operations of dilation, closing and pixel bridging were used. The final step is filling the contour. The results of these operations can be seen in Figure 4.7 and performance metrics are presented in Table 4.2.

Table 4.2: Mean results obtained with Canny's algorithm.

ACC (%)	**SP (%)**	**SE (%)**	**HD (px)**	**XOR (%)**	**Dice (%)**	**Jacc (%)**
85.65	68.17	95.35	144.25	40.62	70.69	62.66

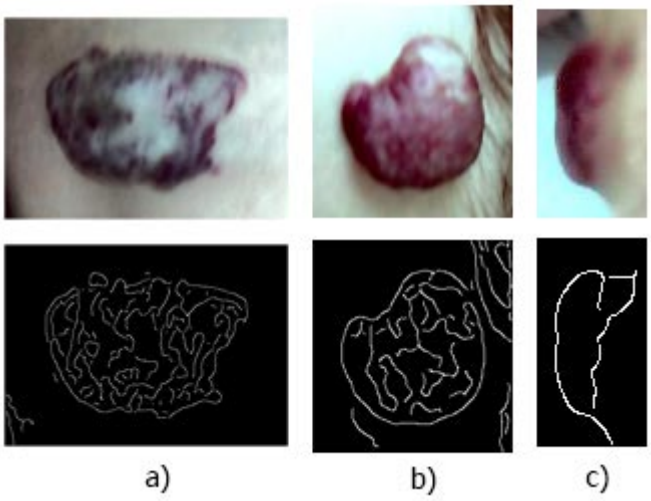

Figure 4.6: Three examples of the lesion's edge detection using Canny's algorithm.

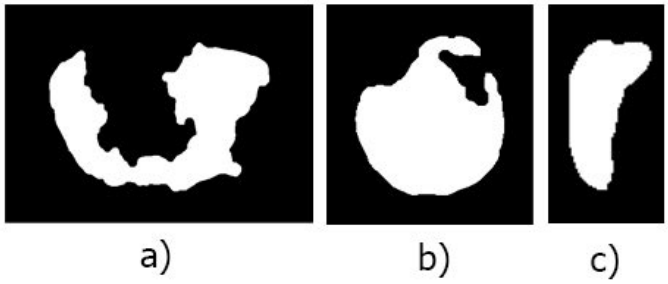

Figure 4.7: Segmentation results after postprocessing.

A major problem of this method is to obtain a fully closed contour or disconnected edges that can be easily connected in the postprocessing steps. If the contrast between the lesion's border and the skin is low then the edges are hard to detect and the postprocessing methods can not obtain a fully closed contour. This is a significant drawback in the present scenario, with high variability in the acquired data. In Figure 4.8, are shown some examples of images where the detected contour is very different from the desired result. In 5 cases it was not possible to obtain a closed contour.

4.3 Fuzzy c-Means Clustering

Again, the input image used in this method is the a^* channel of the preprocessed image. The block diagram for this framework is shown in Figure 4.9.

The number of clusters is set to $c = 2$ and the membership values of each pixel intensity is computed using the FCM algorithm. Then, these values are inspected and each pixel is labeled as belonging to the class with the highest corresponding membership, constructing the binary image (Figure 4.10).

As the image's color and contrast have high variability, many pixels are mislabeled. This is more pronounced in images exhibiting regression, as the color of the pixels from those areas starts

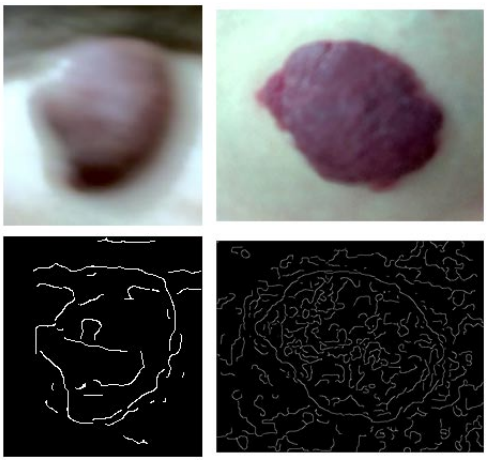

Figure 4.8: Detected edges for two of the bottom results. On the left, the disconnected edges do not become connected and on the right the method over-detects edges that do not belong to the lesion's border.

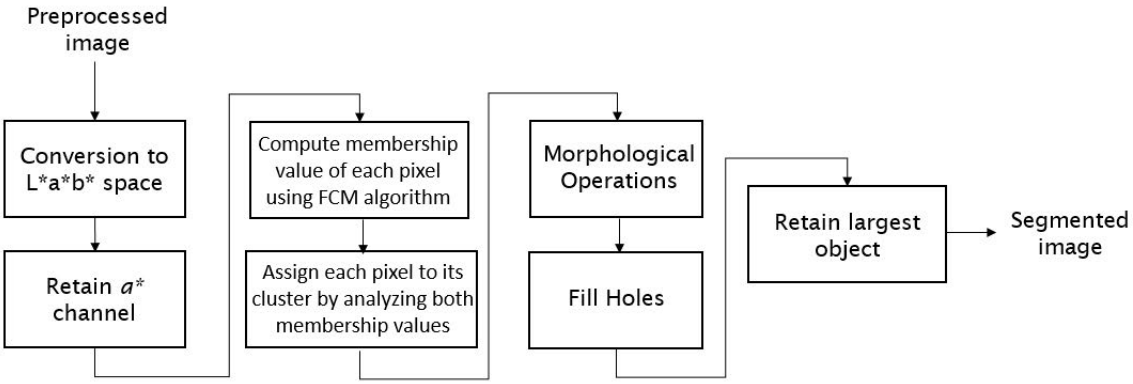

Figure 4.9: Block diagram of the framework devised to use FCM clustering.

fading from red to gray. Therefore, some postprocessing operations of open, erosion and hole filling are applied. The final segmented image is obtained by preserving the biggest object in the image, as shown in Figure 4.11. Table 4.3 displays the performance metrics obtained for this method.

Table 4.3: Mean results obtained with FCM.

ACC (%)	**SP (%)**	**SE (%)**	**HD (px)**	**XOR (%)**	**Dice (%)**	**Jacc (%)**
84.44	89.38	81.88	138.42	44.86	81.08	70.45

Similar to the results using Otsu's method, in some images the postprocessing methods are not enough to eliminate areas that do not belong to the lesion, or to bridge closely disconnected pixels to form a more accurate segmentation of the lesion. Some examples of images with poor segmentation performance are shown in Figure 4.12.

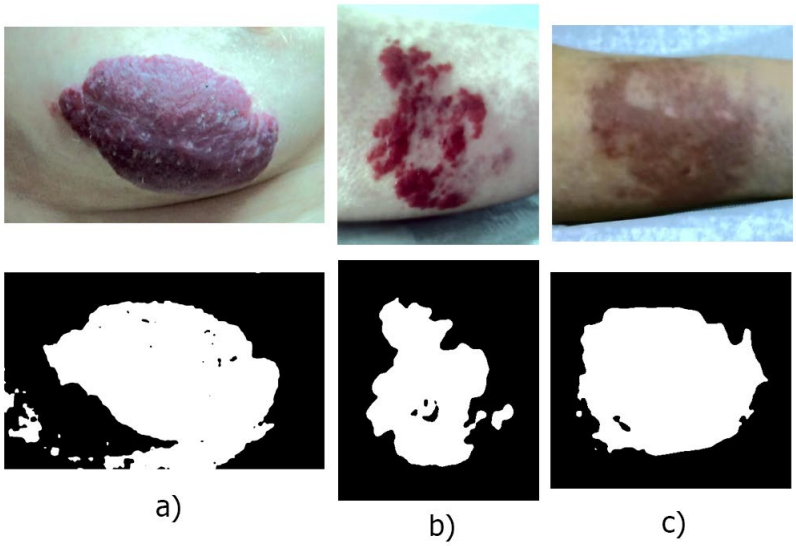

Figure 4.10: Results obtained with FCM clustering.

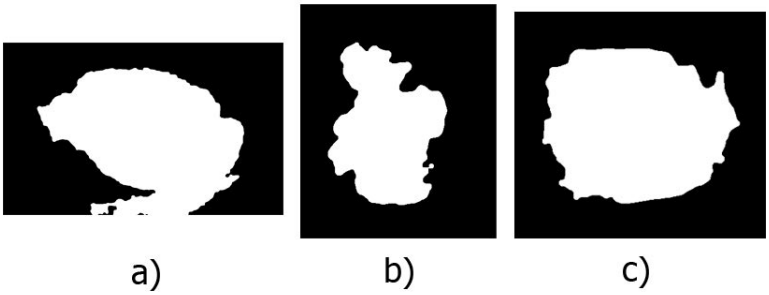

Figure 4.11: Results obtained after postprocessing

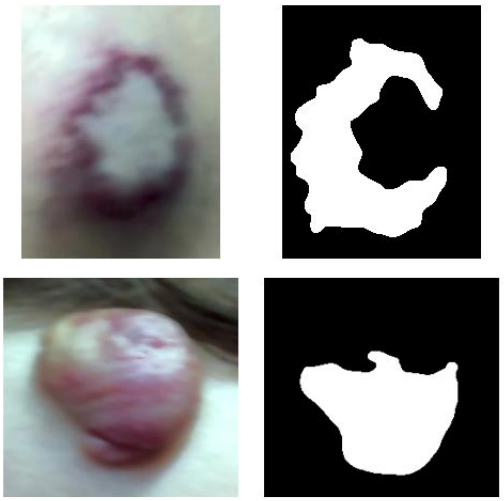

Figure 4.12: Examples of poorly segmented results.

4.4 Boundary Detection with Gradient Vector Flow Snakes

This method requires the input image - the a^* channel - and the snake initialization. Given the input image the edge map is computed, from which the gradient vector flow field is obtained. The number of iterations for its calculation was set at 80 and the regularization parameter at $\mu = 0.2$, which were empirically defined. The block diagram for the presented method is shown in Figure 4.13.

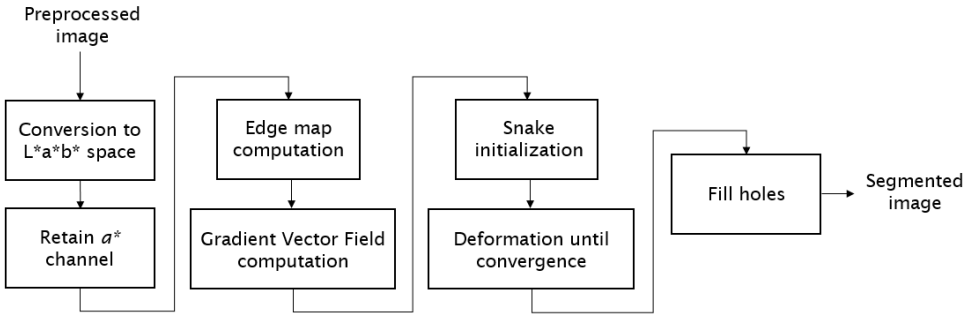

Figure 4.13: Block diagram of the framework devised for using GVF snakes.

In Figure 4.14 one can see the input image, the corresponding edge map and the computed GVF field representation. The snake initialization was set to be an ellipse centered at the image centre, with parametric equations $x = x_c + 0.6 l_w cos(t)$ and $y = y_c + 0.7 l_h sin(t)$ where l_w is half the image's width and l_h is half the image's height. This initialization can be seen in Figure 4.15, as well as the deformation at 3 different iterations. Results obtained with this method are presented in Table 4.4, where the number of iterations for snake deformation was 25.

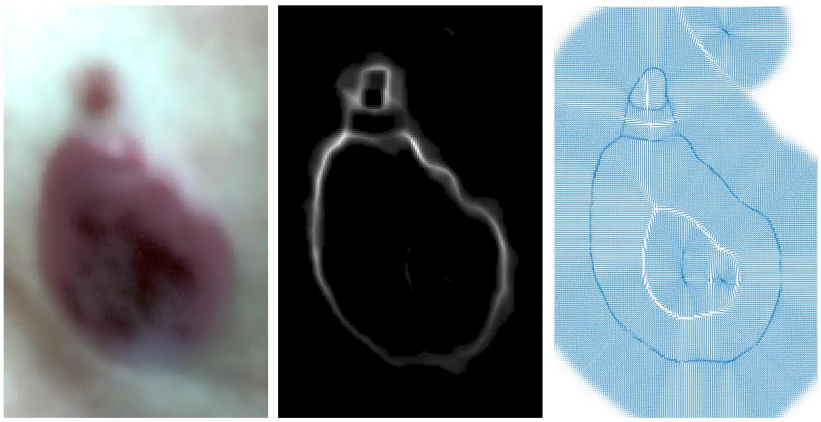

Figure 4.14: Example of input image (a), corresponding edge map (b) and gradient vector flow field (c).

One of the main drawbacks from this method is the computation of the edge map. If the edges are not well defined, then the GVF will not allow for a good deformation of the snake, resulting in poorly segmentation results. Such an example is presented in Figure 4.16, where the edges on the

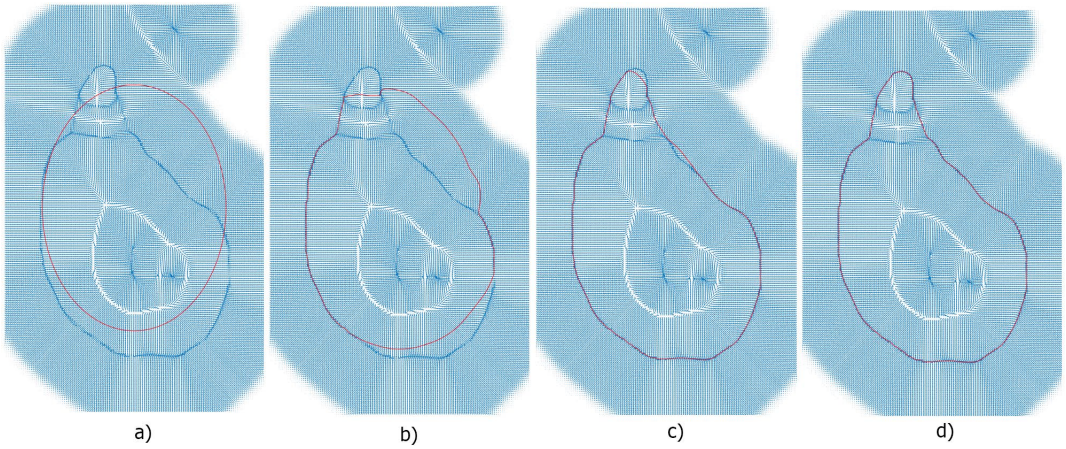

Figure 4.15: Snake initialization (a), followed by successive iterations, $i = 5$ (b), $i = 10$ (c) and $i = 25$ (d).

Table 4.4: Mean results obtained with GVF snakes.

ACC (%)	**SP (%)**	**SE (%)**	**HD (px)**	**XOR (%)**	**Dice (%)**	**Jacc (%)**
91.83	84.87	95.89	97.90	23.52	87.44	78.56

left and right are not well defined (top), resulting in a poor gradient field at those locations, and subsequent poor segmentation (bottom).

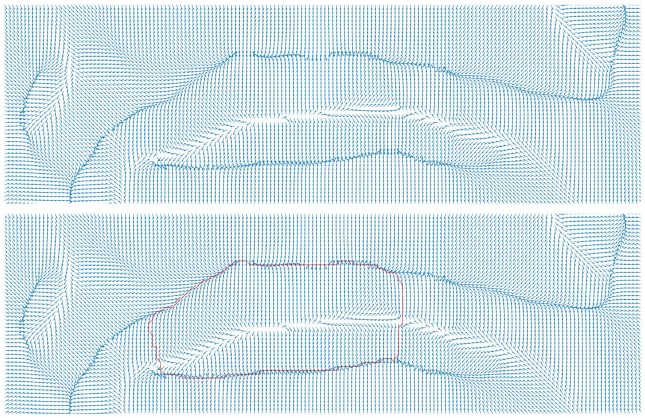

Figure 4.16: Examples of a poorly defined GVF that results in a suboptimal snake convergence.

One straightforward solution would be to use a sharpening filter. However, this would highlight spurious edge pixels, that would result in convergence points in the GVF field that hinder the snake deformation to an accurate result.

The snake initialization is also an important factor. The principle was to make it as automatic as possible, but at a position where it would not be affected by gradient flows pointing away from the edge. In Figure 4.17 are shown two different bad initializations for the same lesion and their convergence after 25 iterations.

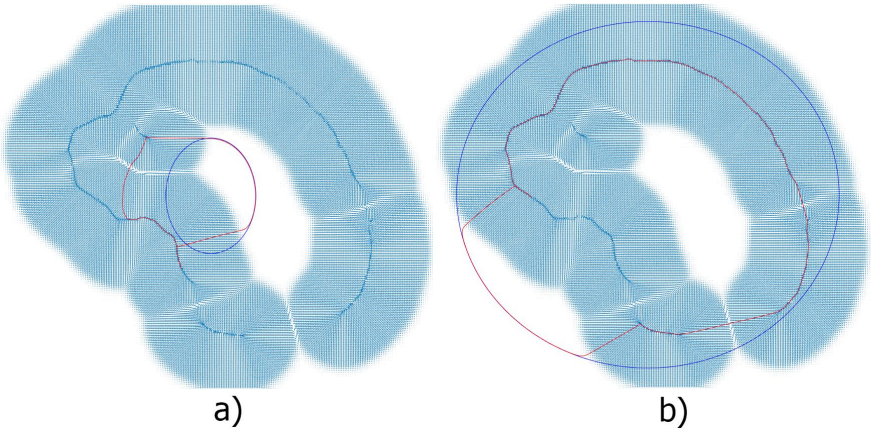

Figure 4.17: Example of two different snake initializations (green) that produce very distinct results (red) due to the GVF not being well defined in some areas. In image (a) the snake is initialized very close to the center of the image and in image (b) is initialized far way from the center.

4.5 Segmentation with Closed Shortest Path

The last of the "shallow" methodologies tested was the closed shortest path calculation. The input is still the *a** channel, and the seed point is located at the image's "center of mass". This center of mass is simply the weighted average of the pixels along the rows and the columns of the image. The block diagram of Figure 4.18 provides the detailed steps for this method.

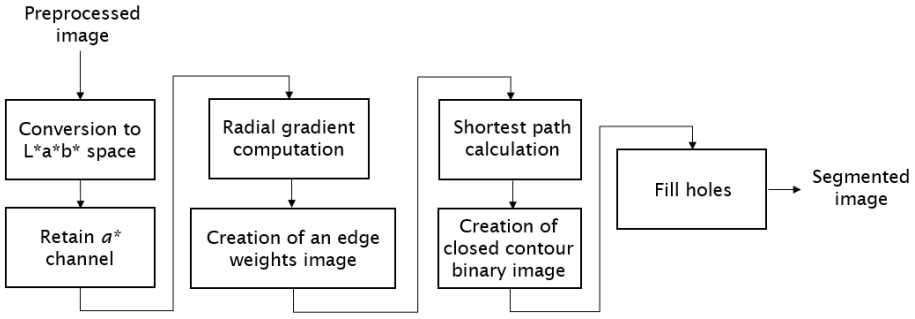

Figure 4.18: Block diagram of the framework devised for using the shortest path calculation.

A kernel of size dependent on the input image's is defined to compute the gradient along the radial directions, and the weight of each pixel is then calculated (Figure 4.19b). This weight, along with the distance to the seed point are used to compute the cost of each edge and to find the shortest path. In Figure 4.20 can be seen the results considering 8 neighbors (4.20b) and 20 (4.20c).

The number of neighbors that yielded the best results was 20, which produces smoother edges when compared to considering 8 neighbors. Obtained performance metrics for this method are presented in Table 4.5.

The main disadvantage of this method is that it cannot properly find the closed contour of objects that have a strong inwards inflection, like the example of Figure 4.21. Nevertheless, this

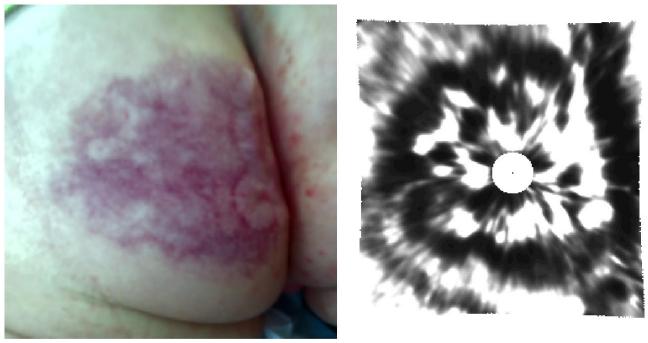

Figure 4.19: Input image (a) and respective radial derivative image (b).

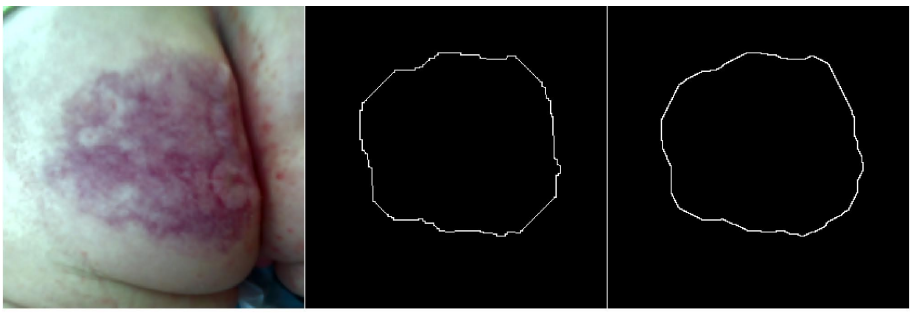

Figure 4.20: Examples of segmentation of input image (a) using 8 neighbors (b) and 20 neighbors (c).

Table 4.5: Mean results obtained with the Shortest Path calculation.

ACC (%)	SP (%)	SE (%)	HD (px)	XOR (%)	Dice (%)	Jacc (%)
93.37	91.79	94.28	50.13	19.14	90.35	83.04

method presented the best results of all the experimented shallow methodologies. The major reasons for this from our point of view are that: it always generates a closed contour, eliminating the need for postprocessing to connect edges; the gradients are computed radially, with an image size dependent kernel, resulting in a wide area of responses where the real edges may be located; even if the real contour has a higher cost than the computed contour, one can assume that the edges considered are always very close to the real ones, given the difference in the Haussdorf distance when compared to other methods.

4.6 Segmentation using Convolutional Neural Networks

The approach chosen for this method was to use squared patches of the image, centered at a given pixel, as input. This central pixel is classified either as hemangioma or non-hemangioma. The network is therefore learning to identify to which class a pixel belongs to, given its surroundings. An equal number of patches are extracted randomly for both classes. These patches are created

Figure 4.21: Example of a lesion with an inward inflection (a). In (b) it can be seen that the method has difficulty in dealing with these types of contours (ground-truth border presented in white and obtained border in red).

with dimensions depending on the largest side of the image, to ensure that its resolution is proportional to the image's resolution. The resolutions of the patches were chosen as $1/10$ and $1/5$ of the image's largest side as shown in Figure 4.22.

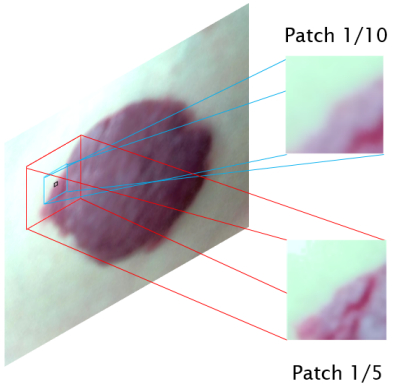

Figure 4.22: For a given image, a central pixel is defined and two possible patches are drawn.

Two different architectures were considered in this work. One is a standard linear CNN (Figure 4.23) and the other follows a DAG topology that combines two of these linear architectures (Figure 4.24). We will call the single input architecture 1R-CNN, and the double input architecture 2R-CNN (R standing for Resolution, referring to the resolution of the patches). The configuration of each layer is represented in Table 4.6, and in the CNNs representations of Figures 4.23 and 4.24 are shown the dimensions of the data as it goes through the network.

Table 4.6: Configurations for each layer composing the CNN architectures.

Layer	0	1	2	3	4	5	6	7	8	9
Type	Input	Convolutional	Pooling	ReLU	Convolutional	Pooling	ReLU	Fully Connected	Fully Connected	Loss
Name	*Input*	*Conv1*	*Pool1*	-	*Conv2*	*Pool2*	-	*FC1*	*FC2*	-
Kernel Size	-	6	2	1	5	3	1	3	1	1
Number of Filters	-	60	-	-	60	-	-	500	2	-
Stride	-	1	2	1	1	3	1	1	1	1
Padding	0	0	0	0	0	0	0	0	0	0

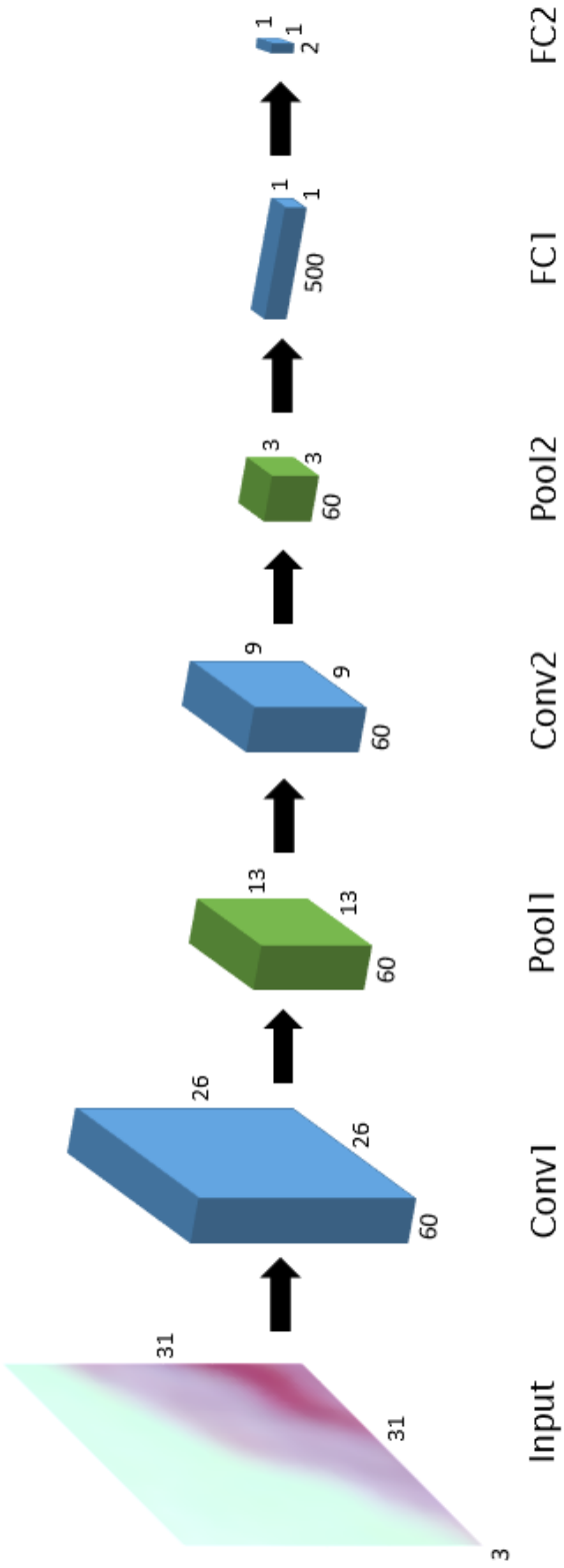

Figure 4.23: Linear CNN architecture, 1R-CNN, with one input.

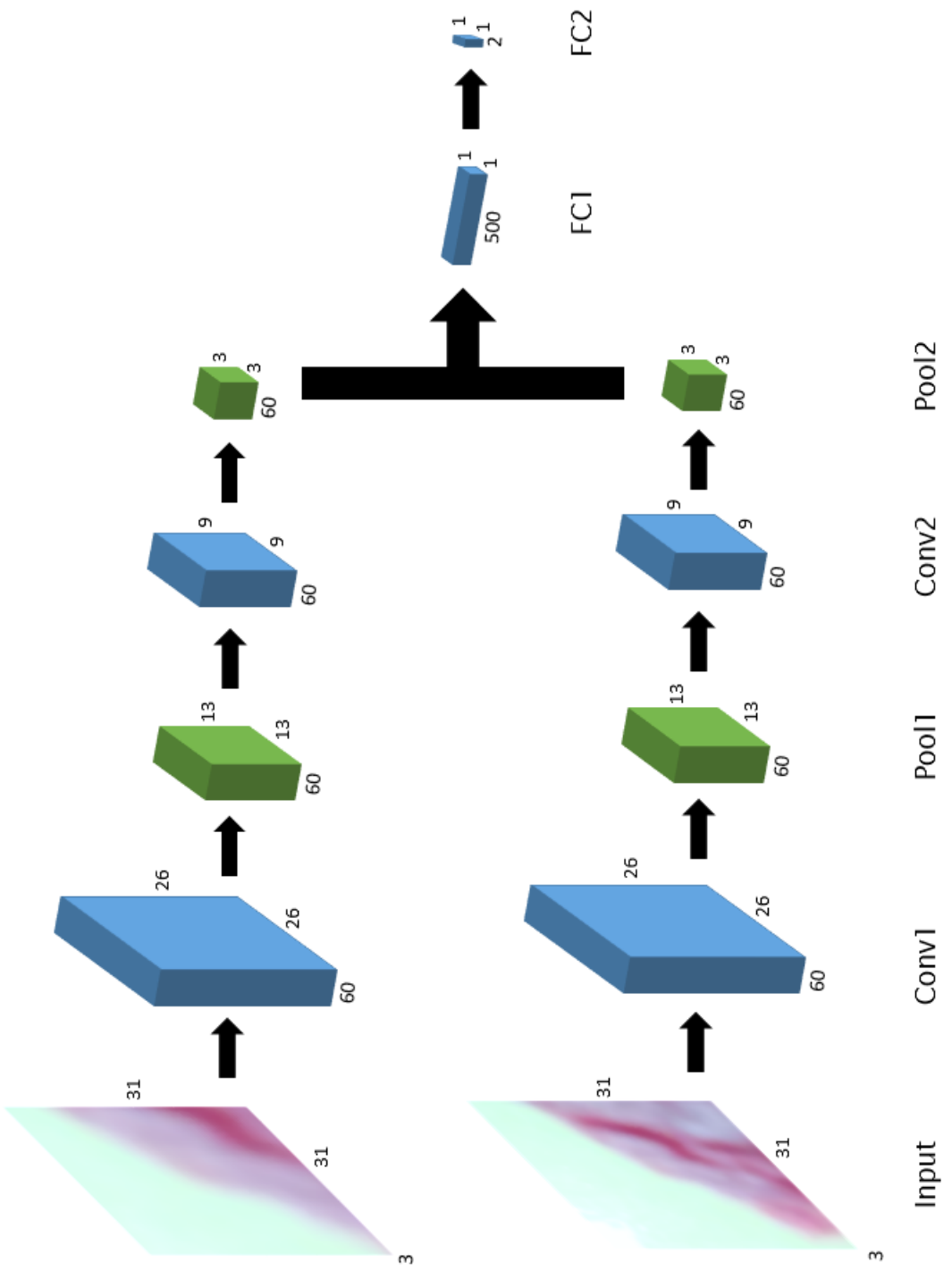

Figure 4.24: DAG CNN architecture, 2R-CNN, showing the two different inputs.

For both approaches the weights and bias were randomly initialized, the learning rate set to 10^{-4}, and the number of epochs set to 1000. For the implementation of the networks the MatConvNet toolbox for MATLAB was used, and training was carried out using a NVIDIA GTX960M.

The tested networks are presented in Table 4.7 along with their inputs and training and validation errors. The error evolution along the epochs for these networks is presented in Figure 4.25. The 200 images were randomly split into 4 subsets of 50 images. Two of these sets were selected for training and one for validation. From each training image 750 patches belonging to the each class were extracted, with a total number of 150000 patches for training and 75000 patches for validation. On both architectures the patches are all resized to a size of $31 \times 31 \times 3$ to which the mean patch values of the training data are subtracted.

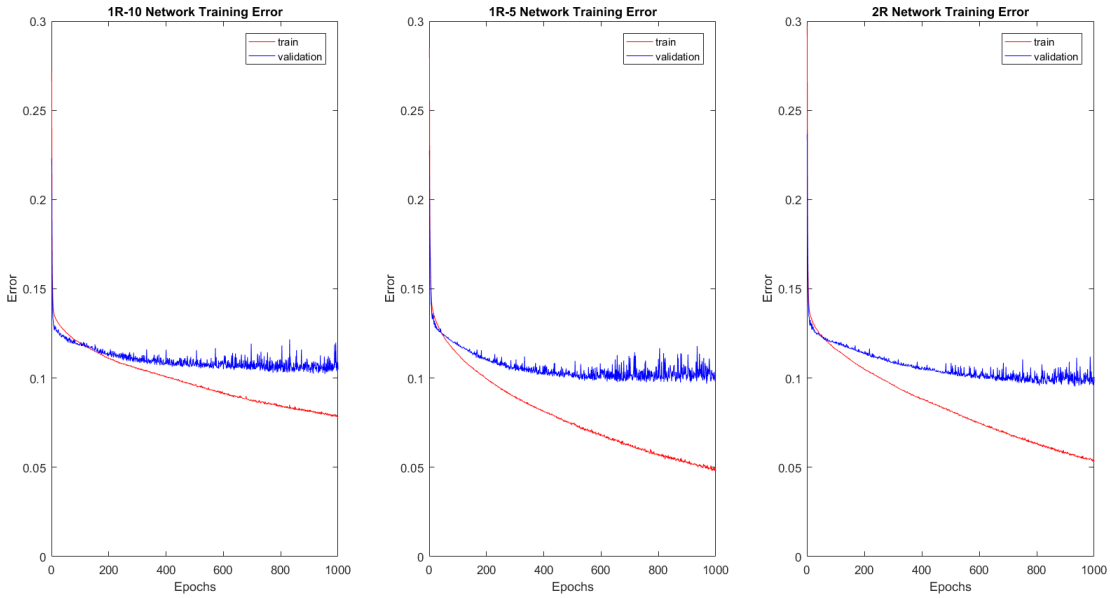

Figure 4.25: Training evolution for the tested network architectures.

Table 4.7: Tested architectures, their respective inputs and training and validation errors.

CNN	**Patch Resolution**	**Train Error (%)**	**Validation Error (%)**
1R-10	1/10	7.87	10.49
1R-5	1/5	4.87	10.43
2R	1/10 and 1/5	5.38	10.07

After training the networks, the segmentation of the images is made following the block diagram of Figure 4.26. For each pixel of the image a patch is extracted, resized and evaluated through the network, producing a vector of probabilities with size 1×2, with a score for each class. The image's corresponding pixel is labeled (0 for non-hemangioma and 1 for hemangioma) to the class with highest probability. In Table 4.8 are shown the performance results of the various CNNs used in this approach and in Figure 4.27 are some segmentation examples after postprocessing.

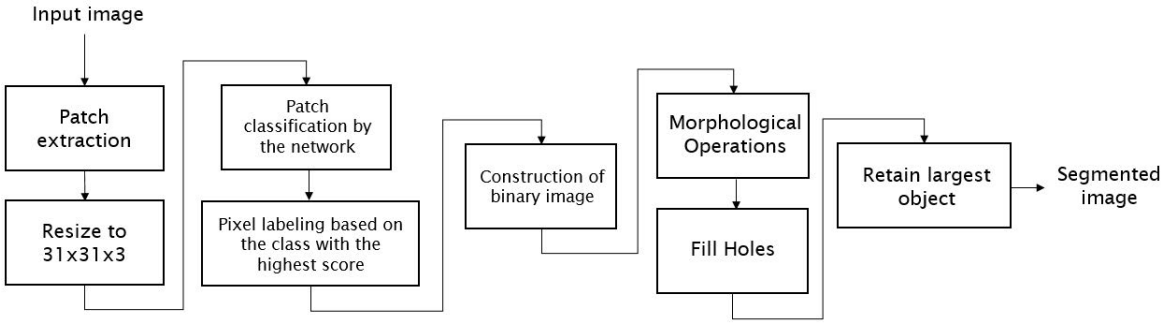

Figure 4.26: Block diagram devised for the segmentation using the trained CNNs.

Figure 4.27: Some examples of results obtained using the trained CNNs.

Table 4.8: Mean results obtained for the various CNN architectures.

CNN	**ACC (%)**	**SP (%)**	**SE (%)**	**HD (px)**	**XOR (%)**	**Dice (%)**	**Jacc (%)**
1R-10	91.14	92.71	90.53	118.20	25.63	88.3	79.99
1R-5	90.16	81.57	95.14	119.53	27.86	84.50	75.01
2R	92.99	91.67	93.92	105.56	20.10	90.20	82.95

4.7 Segmentation using Fine-Tuned Convolutional Neural Networks

Regarding the Transfer Learning approach for segmentation, the same patch-wise training was performed. To this end, the publicly available SMARTSKINS melanoma dataset (Vasconcelos et al., 2014) was used, with two examples presented in Figure 4.28. First, a set of 80 melanoma images taken with a smartphone was randomly divided into 4 subsets of 20 images each. The images were preprocessed in the same fashion as the hemangioma images. Since the goal was not the segmentation of melanoma lesions, but rather to transfer the network's weight, and since there were few images in this dataset, three subsets were used for training and one for validation. From each image, 1250 patches of each class were extracted, in a total of 150000 patches used for training and 50000 patches used for validation. A network using the 1R-10 architecture was trained using these patches, in the same fashion as explained in the previous subsection. Then, fine-tuning of the last fully connected layer was performed using the hemangioma data, "freezing" the parameters of the other layers . Both training plots are presented in Figure 4.29, and segmentation performance results obtained using this fine-tuned network are presented in Table 4.9.

The goal of the previous approach was to transfer information from different lesions that were acquired with a similar sensor (a smartphone). However, few melanoma data was available, which

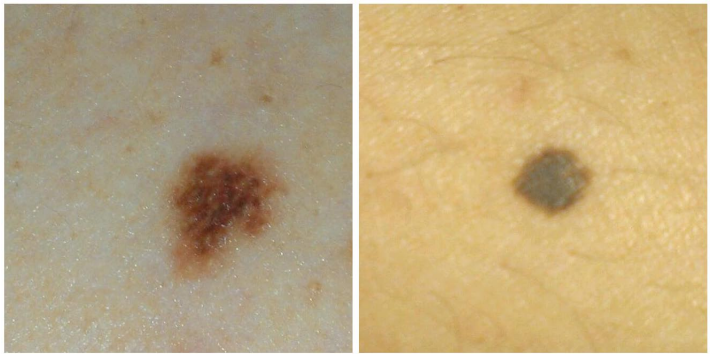

Figure 4.28: Two examples of images of melanoma from the SMARTSKINS dataset.

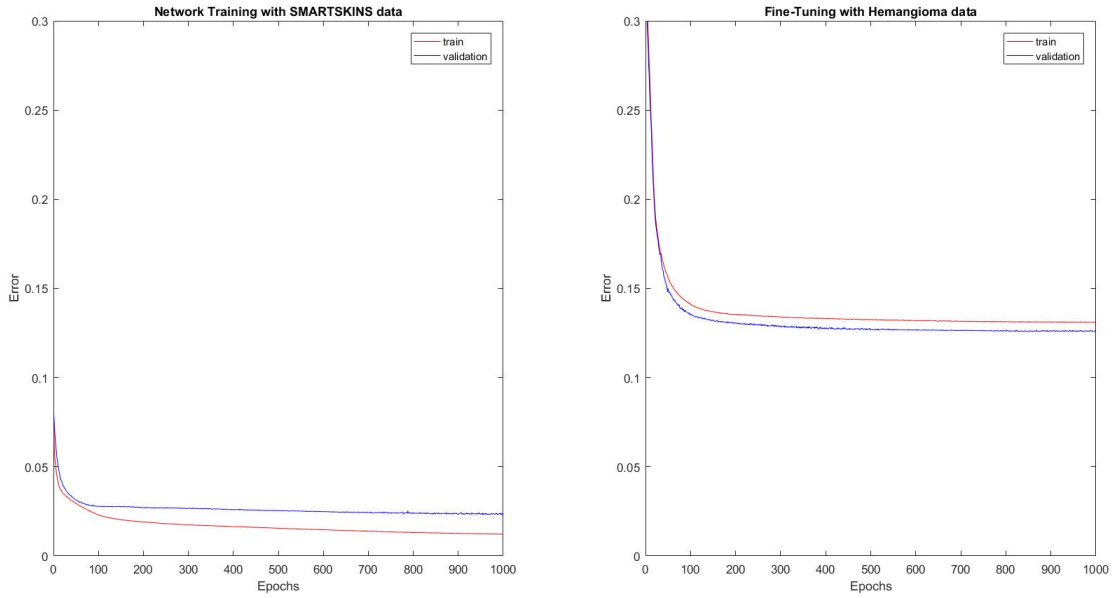

Figure 4.29: Training plots for the single-stage transfer learning approach.

Table 4.9: Mean Performance results obtained for segmentation using the single-stage transfer learning approach.

CNN	**ACC (%)**	**SP (%)**	**SE (%)**	**HD (px)**	**XOR (%)**	**Dice (%)**	**Jacc (%)**
1R-10 1 stage	87.84	88.57	87.85	128.23	35.26	83.47	73.14

poses a problem when using CNNs, that require large volumes of data for training robustly without overfitting. There are numerous publicly available datasets of melanoma, but in those the images were acquired with a dermoscope.

Dermoscope images (Figure 4.30) are very different from those acquired with a smartphone: the lesion is looked at through a magnifier that has a non-polarizing light source and which may need a medium between the lens and the skin. Therefore, to make avail of the large datasets available, a two stage transfer learning framework was devised, as shown in Figure 4.31. First the network is trained using the dermoscope images and the final parameters are transferred to become

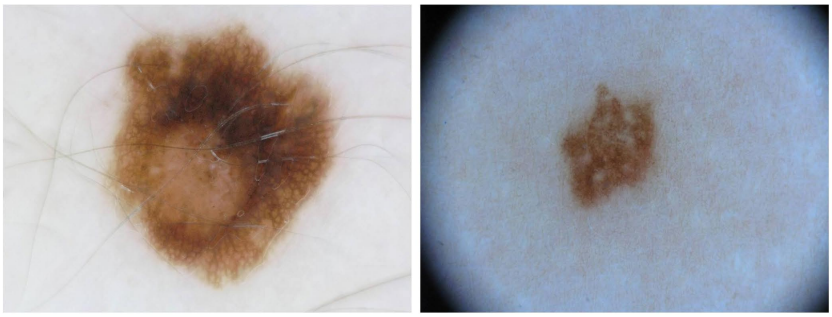

Figure 4.30: Two examples of dermoscope melanoma lesions from the ISIC dataset.

the next CNN initialization. Then, this network is fine-tuned using the melanoma images acquired with smartphone - cross-sensor transfer learning. This fine-tuned network is then fine-tuned again using the hemangioma images - domain transfer learning.

The melanoma dataset provided in the ISIC 2017 challenge (Berseth, 2017) was used to train the network for the first transfer learning stage. This dataset has 2000 dermoscope melanoma images and their respective ground-truth segmentations. The images were preprocessed as explained before and, since they can often contain hairs, a simple hair removal algorithm based on the one of Jaworek-Korjakowska and Tadeusiewicz (2013) was used. The image set was divided in 1500 images for training and 500 images for validation, in which 300 patches were extracted from each image. Again, a network following the 1R-10 architecture was chosen for training. Then, fine-tuning of the last two fully connected layers is performed using the data from the SMARTSKINS dataset. Finally, fine-tuning of the last fully connected layer is performed using the hemangioma data. The training and validation error plots for this two stages are shown in Figure 4.32. Performance results obtained with this framework are presented in Table 4.10 and some segmentation examples are presented in Figure 4.33.

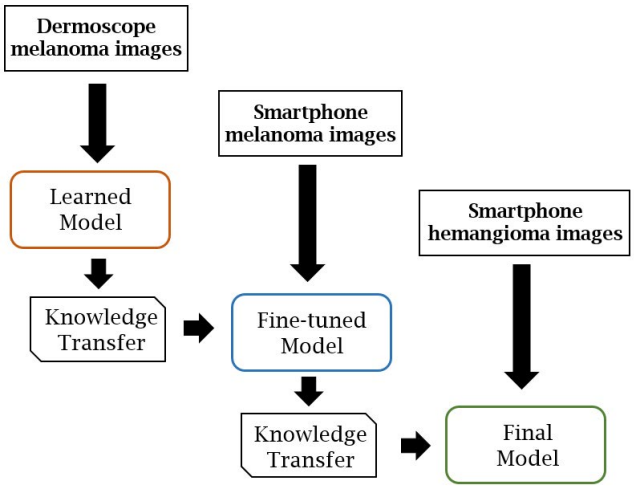

Figure 4.31: Two stage transfer learning framework devised for this work.

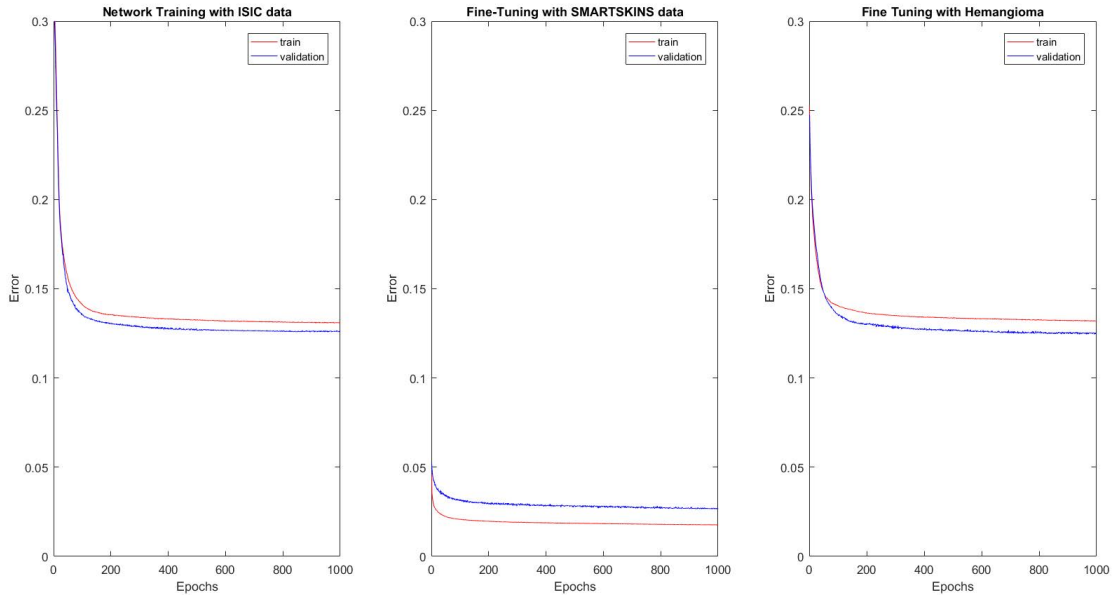

Figure 4.32: Training plots for the two-stage transfer learning approach.

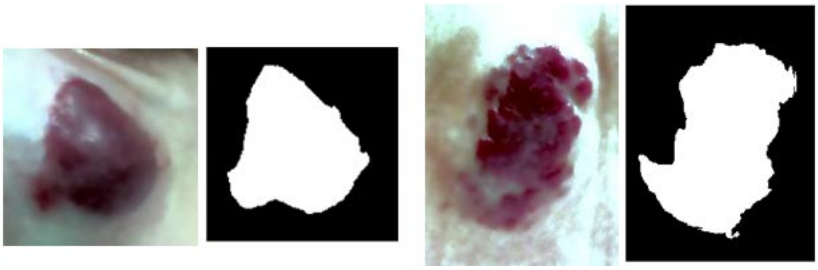

Figure 4.33: Some examples of results obtained using the fine-tuned CNNs.

4.8 Comparison of Results

The key focus of the present thesis was the segmentation of the hemangioma area. This chapter presented the results obtained for each considered approach detailed in Chapter 3, with their strong and weak points discussed based on their performance. Here, we present a compiled version of these results (Table 4.11) and compare them to each other and with those reported in the state-of-the-art, ending with conclusions that are drawn from this analysis.

Table 4.10: Mean performance results obtained with the two-stage transfer learning framework.

CNN	**ACC (%)**	**SP (%)**	**SE (%)**	**HD (px)**	**XOR (%)**	**Dice (%)**	**Jacc (%)**
1R-10 2 stages	87.94	87.46	88.64	127.38	34.91	83.30	73.05

Table 4.11: Compiled results obtained for all the experimented methods.

Method/ CNN	**ACC (%)**	**SP (%)**	**SE (%)**	**HD (px)**	**XOR (%)**	**Dice (%)**	**Jacc (%)**
Shallow Methodologies							
Otsu	84.30	89.58	81.57	138.92	45.26	81.05	70.4
Canny	85.65	68.17	95.35	144.25	40.62	70.69	62.66
FCM	84.44	89.38	81.88	138.43	44.86	81.08	70.45
GVF	91.83	84.87	95.89	97.90	23.52	87.44	78.56
SP	**93.37**	**91.79**	**94.28**	**50.13**	**19.14**	**90.35**	**83.04**
Deep Learning							
1R-10	91.14	92.71	90.53	118.20	25.63	88.25	79.99
1R-5	90.16	81.57	95.14	119.53	27.86	84.50	75.01
2R	**92.99**	**91.67**	**93.92**	**105.56**	**20.10**	**90.15**	**82.95**
Transfer Learning							
1R-10 1 Stage	87.84	88.57	87.85	128.23	35.26	83.47	73.14
1R-10 2 Stages	87.94	87.46	88.64	940.17	34.91	83.30	73.05

4.8.1 Shallow Methodologies

Focusing on the shallow methodologies, the calculation of the shortest path to obtain a closed contour is the one that allows for better results. GVF snakes are right behind in terms of performance. On the other end of the performance ranking is Canny's algorithm, with the worst results. All these three methods are edge-related, which indicates that, as expected, the borders of the hemangiomas are an important feature to take into account for segmentation. So why this discrepancy in performance?

The top two referred methodologies have one thing in common: some form of initialization. In one, the seed point is initialized at the center of mass of the object, which provides a good estimation on where the hemangioma is located and ensures that the final contour will enclose that area; in the other method, the closed contour (snake) is initialized, removing the need for postprocessing methods to obtain a fully closed border. We can view these two methods as some form of refinement of the contour, where either we refine a full initialization or we refine the contour step by step, choosing the "best" edge out of a set weighted by their radial gradient - more precise methods. In Canny's algorithm, all possible edges on the image are found, including those outside of the lesion's border, edges that are part of the lesion's border and edges that may be inside of the lesion - a less precise method. This method is also very dependent on the quality of the image: low contrast diminishes the number of detected edges and those detected may be disconnected from each other; noise present in the image gives rise to detected edges that do not belong to the lesion. So, the key point here is that edge-based methods that have a prior knowledge about the contour or the area where it could be located present an advantage over the other tested edge-based approaches.

In the middle of the performance ranking we have the clustering methods: thresholding by

Otsu's method and fuzzy c-means clustering. These two methods are based on the intensity of the pixels of the image, and present very similar results. The reason these fall behind compared to the methods referred above is relatively straightforward. In both we make use of prior knowledge about the red color of the hemangioma, which will be the separability criterion: pixels with more intensity towards red colors will be clustered together, hopefully all representing the lesion. This assumption has some shortcomings: in some areas of the lesion, specially around the border, pixels may have low contrast, resulting in over segmentation; in some areas of the skin, the pixel intensity may be similar to the intensity of the lesion, resulting in the assignment of those pixels to the wrong cluster; finally, and one of the most significant drawbacks, is that, in lesions exhibiting involution, the involuted areas will not be clustered with the lesion, giving that the color and texture of these pixels has changed considerably, and their intensities no longer resemble those of the original lesion. The reader may be impelled to think that a simple solution would be considering three clusters (or two thresholds, regarding Otsu's method) instead of only two. However, this would hinder performance in images that do not show involution, and even with involuted lesions, the third cluster could be labeling background pixels instead of the intended regressed pixels. Also, to get the final two-class segmentation, some form of additional step in the postprocessing stage would be necessary.

The results obtained with the shortest path calculation also surpass those reported in the state-of-the-art. It is important to note that the datasets used in the state-of-the-art and the one used in this work are very different: none used smarthphone photos, but rather digital cameras with better quality; they all had fewer images available. Given that, we can then conclude that our results are robust and meaningful, also backed up by the variability that the images have.

4.8.2 Deep Learning Methodologies

After establishing and examining the baseline performance with the shallow methods, let us turn our discussion to the performance obtained with deep learning methodologies.

As one can see from Table 4.8, these results surpass most of those obtained in the previous section and in the state-of-the-art. By comparing them to each other, one can draw some interesting conclusions: for the same network, patches that consider a smaller area around the central pixel, and therefore have a higher resolution of that area, perform better than those with lower resolution around the central pixel. This is somewhat intuitive, as one can expect that, with a higher level of detail, the network could "learn better" how to discern the pixel's class. Another interesting aspect is the improvement in performance with the use of two different inputs, as seen by the results obtained with the 2R-CNN architecture. This implies that, in the cases that the network had troubling deciding correctly with only one patch, information from the other patch, which has another "point of view" helped improve this decision.

The results using CNNs were a bit unexpected, in a good way. The major findings in the area of CNNs report networks trained with tens of thousands or even millions of images. In the available dataset we have only 200 images, which would make training a plausible network impossible, even if classical data augmentation techniques were considered. The patch-based

approach allows for this training without having to transform the available images, and raises the idea of training a CNN to a more reasonable level. Even so and despite the volume and variability of the dataset and the depth of the network, this approach still achieved very good results. The need for the postprocessing step to eliminate incorrectly classified pixels suggests that the filters of the network are leaning towards learning the parameters based on color and/or texture features of the patches. As we have concluded with the results from shallow methodologies, segmentation of hemangiomas based on color features has its drawbacks.

The performance of the best deep methodology and best shallow methodology are very close, with the computation of the shortest path slightly surpassing the 2R-CNN. However, if we look at the Haussdorf distance obtained for both methods, the shortest path calculation greatly outperforms the CNN. This comes to corroborate the reasoning laid out throughout this discussion: in the shortest path computation an edge will always be chosen so that it belongs to a contour enclosing the seed point, which means that, even if the edge chosen to be added to the path is not the one that belongs to the lesion boundary, it will not be very far off from the real one (excluding lesions with strong inwards inflections, as shown in section 4.5). This explains the differences in the obtained Haussdorf distances and further supports the rationale that edge information and prior knowledge about the lesion's surroundings greatly influence performance.

4.8.3 Transfer Learning

Concerning the single-stage transfer learning approach, results suggest that, as expected, the pre-trained melanoma network can be adapted to IH segmentation, but with some difficulties. The segmentation performance surpasses most of the shallow methodologies, but fall short to the results obtained with the other CNNs. Despite being possible to transfer some knowledge from one task to the other, in the end, the data is still very different and the network's training error cannot converge to a smaller value. Nonetheless, networks such as this can be used to perform some form of raw segmentation that is then refined using some other approach.

In the two-stage transfer learning approach, one can draw interesting conclusions. First, the knowledge transfer between acquisition sensors is proven to be a success, as the network's training error is very low. This is somewhat expected: if we look at the difference between the dermoscope melanoma images and smartphone melanoma images used, we can see that the dermoscope images have much higher detail and heterogeneity and so, refining the network to images with lower resolution (from a lesion detail point-of-view) will therefore be easy. The same cannot be said for the knowledge transfer between domains. Analyzing the training plots for the final fine-tuning stage and subsequent hemangioma segmentation results, we can conclude that, despite it's good training results, the first fine-tuning stage does not have a significant impact on the transfer of knowledge from melanoma and hemangioma, as the obtained performance is very similar to the single-stage transfer learning approach.

In summary:

- The computation of the shortest path to obtain a closed contour enclosing a seed point, on which we have prior knowledge that it may be the center of the lesion, revealed itself as the best methodology;
- Despite the red color being a prominent feature of hemangiomas that has been a focus point in many works, taking into account edge information remains a solid approach, despite not having characteristics exclusive to hemangiomas;
- CNNs proved to be a powerful tool for segmentation, achieving good results despite the small amount of data available and the depth of the network. The use of more intricate architectures may help achieve higher performances.
- Transfer learning results were not as good as ambitioned, but can still outperform most of the shallow methodologies. Knowledge transfer between sensor types proved to be better than between lesion types, but has little impact on the results for the two stage transfer learning framework.

In the next and final chapter, the main conclusions about this thesis are outlined, as well as improvements that can be done to the explored methodologies and future lines of work that can be pursued to improve findings and research in the area of CAD systems for infantile hemangiomas.

Chapter 5

Conclusions and Future Work

This final chapter will act as both a summary of the main conclusions attained as result of the scientific contributions presented, as well as the roadmap to future lines of research in the area of infantile hemangiomas.

5.1 Conclusions

The presented work has brought a contemporary vision on the problematic of infantile hemangiomas and has explored new paths that have not yet been considered in the state-of-the-art. The main conclusions on each topic explored are as follows:

- **Dataset creation:** Starting by the development of the dataset, a wide number of photos acquired unconstrainedly of various hemangioma cases and in various consultations were collected. These were manually segmented, giving rise to the ground-truth images. From these photos, a ROI was considered by expanding the bounding box enclosing the lesion by 20%. A total of 200 images from 29 patients, on which 13 have photos from more than one consultation is then assembled. This dataset can be useful to many lines of research: hemangioma classification, risk assessment, lesion segmentation, lesion registration and lesion evaluation (either on a single occasion or over time). Also, the unconstrained setting allows for the development of more robust methodologies. To the extent of our knowledge, this is the biggest dataset of the kind created so far, both in number of images and number of patients. An improvement would be adding some form of reference scale to the images, to allow for measurements in real life metrics. Exploring further modalities of acquisition, such as infrared thermography, depth images and ultrasound, would expand the range of applications and provide further sources of information for the various topics of research.

- **Shallow Methodologies:** The focus of the work was the segmentation of the hemangioma lesion. On a first approach and inspired by previous works on the area, several shallow methodologies were tested: Otsu's method for thresholding, Canny's algorithm for detecting edges, fuzzy c-means clustering, gradient vector flow snakes and shortest path computation.

GVF snakes and shortest path calculation obtained state-of-the-art performances, with the latter obtaining the best results. Both methods are based on the edges of the hemangioma and surpass the tested methods that only consider color features. This was attributed to the fact that color is a feature with high variability between patients and that lesions exhibiting regression fail to have a good segmentation with these methods. Exploring some form of fusion of methods and/or features could help improve performance, especially for cases with regression or low border contrast.

- **Deep Methodologies:** After establishing a baseline for the segmentation results, the use of deep learning methodologies arose as a good approach, since it has been in vogue in the last years, with excellent results. Therefore, two convolutional neural network architectures were devised and trained following a patch-wise approach. These patches are extracted from the input images at two different resolutions and fed to the network. Obtained results surpass the state-of-the-art and show that training with patches with better resolution of the area around the central pixel can help obtain a better performance. A network that receives two different patches as input was also developed. The use of two resolutions improved the segmentation results, as the network has more information regarding the area around the central pixel. The performance results obtained with this network were very similar to those obtained with the shortest path method, except in the Haussdorf distance, where it was inferior. This suggests that the network is learning with more regard for color features, rather than spatial/edge information, and corroborates the conclusions drawn on the shallow methodologies that edge information is relevant in this task and that ensemble of methodologies should be a research topic to pursue.

- **Transfer Learning:** Since transfer of learning from pre-trained models has been a relatively new and trendy topic in the machine learning community, the idea of applying it in this dissertation arose. The research on melanoma segmentation is abundant, and so are the number of publicly available images. Therefore, the challenge was to use a pre-trained model on melanoma and fine-tuning it to application on hemangioma images. This was done in two ways: using a set of melanoma images acquired with a smartphone to train a model which was then fine-tuned with IH images; using a set of dermoscopy melanoma images to train model, transfer the knowledge gained and fine-tune that model with the smartphone melanoma set, and fine-tune it again using the IH images. Results showed that this approach does not yield results as good as ambitioned, but can still surpass most of those obtained with the shallow methodologies. Due to the given time frame only one architecture was tested which suggests that results could improve, as the architecture that yielded the best performance in the Deep Methodologies was not experimented. This can be a good starting point for future work.

5.2 Future Work

All that has been discussed opens new pathways to be explored. Future work could start with the development of a detection and risk evaluation system that could classify the hemangioma in one of the four subtypes, identify its phase and evaluate the need for clinical intervention.

Focusing on the segmentation step, there are many possible approaches to try and improve what has already been made. In particularly, the ensemble of different methods and features could help improve performance for lesions with low contrast on the borders or that exhibit regression, by considering different types of features. In the deep learning domain, more intricate architectures could be used, like residual networks or fully convolutional networks, or the use of more input information to help the network to better discern the pixel classification. The problem, when compared with other domains, is the lack of data and the difficulty in obtaining it, given all that has been discussed about acquisition in the medical scenario. The idea of transfer learning has been proven a good route, as there is a lot of research work done in the area of melanoma detection and segmentation. Results have shown that the fine tuning of a solution for a similar task is a good approach, with room for improvement, specially concerning the tested architectures.

Regarding the evaluation stage, little work has been done in the state-of-the-art for what constitutes one of the main goals of the devised CAD system. In particularly, image registration was only focused in the work of Zambanini et al. (2010), and results suggest a clear room for improvement. One main problem in this stage is that the changes in the lesion due to involution might be considerable and thus limiting the ability to identify tie points to align images.

All these suggestions could complement the work developed in this thesis and open way to the development of a full and robust CAD system for the automatic assessment of infantile hemangiomas.

References

Barnard, K., Cardei, V., and Funt, B. (2002). A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *IEEE transactions on Image Processing*, 11(9):972–984.

Berseth, M. (2017). Isic 2017-skin lesion analysis towards melanoma detection. *arXiv preprint arXiv:1703.00523*.

Bezdek, J. C., Ehrlich, R., and Full, W. (1984). Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203.

Boon, L. M., Enjolras, O., and Mulliken, J. B. (1996). Congenital hemangioma: evidence of accelerated involution. *The Journal of pediatrics*, 128(3):329–335.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26.

Burkes, S. A., Patel, M., Adams, D. M., Hammill, A. M., Eaton, K. P., Randall Wickett, R., and Visscher, M. O. (2016). Infantile hemangioma status by dynamic infrared thermography: A preliminary study. *International journal of dermatology*, 55(10).

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698.

Cardoso, J. S., Domingues, I., and Oliveira, H. P. (2015). Closed shortest path in the original coordinates with an application to breast cancer. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(01):1555002.

Cavalcanti, P. G., Scharcanski, J., and Lopes, C. B. (2010). Shading attenuation in human skin color images. In *International Symposium on Visual Computing*, pages 190–198. Springer.

Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., and Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31(6):362–373.

Celebi, M. E., Mendonca, T., and Marques, J. S. (2015). *Dermoscopy image analysis*, volume 10. CRC Press.

Chang, L. C., Haggstrom, A. N., Drolet, B. A., Baselga, E., Chamlin, S. L., Garzon, M. C., Horii, K. A., Lucky, A. W., Mancini, A. J., Metry, D. W., et al. (2008). Growth characteristics of infantile hemangiomas: implications for management. *Pediatrics*, 122(2):360–367.

Chiller, K. G., Passaro, D., and Frieden, I. J. (2002). Hemangiomas of infancy: clinical characteristics, morphologic subtypes, and their relationship to race, ethnicity, and sex. *Archives of dermatology*, 138(12):1567–1576.

Dickison, P., Christou, E., and Wargon, O. (2011). A prospective study of infantile hemangiomas with a focus on incidence and risk factors. *Pediatric dermatology*, 28(6):663–669.

Esterly, N. B. (1995). Cutaneous hemangiomas, vascular stains and malformations, and associated syndromes. *Current problems in dermatology*, 7(3):69–107.

Finlayson, G. D. and Trezzi, E. (2004). Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology.

Finn, M. C., Glowacki, J., and Mulliken, J. B. (1983). Congenital vascular lesions: clinical application of a new classification. *Journal of pediatric surgery*, 18(6):894–900.

Fornaciali, M., Carvalho, M., Bittencourt, F. V., Avila, S., and Valle, E. (2016). Towards automated melanoma screening: Proper computer vision & reliable results. *arXiv preprint arXiv:1604.04024*.

Foster, D. H. (2011). Color constancy. *Vision research*, 51(7):674–700.

Frieden, I. J., Haggstrom, A. N., Drolet, B. A., Mancini, A. J., Friedlander, S. F., Boon, L., Chamlin, S. L., Baselga, E., Garzon, M. C., Nopper, A. J., et al. (2005). Infantile hemangiomas: current knowledge, future directions. proceedings of a research workshop on infantile hemangiomas. *Pediatric dermatology*, 22(5):383–406.

Greenberger, S. and Bischoff, J. (2013). Pathogenesis of infantile haemangioma. *British Journal of Dermatology*, 169(1):12–19.

Haggstrom, A. N., Drolet, B. A., Baselga, E., Chamlin, S. L., Garzon, M. C., Horii, K. A., Lucky, A. W., Mancini, A. J., Metry, D. W., Newell, B., et al. (2006). Prospective study of infantile hemangiomas: clinical characteristics predicting complications and treatment. *Pediatrics*, 118(3):882–887.

He, K., Sun, J., and Tang, X. (2013). Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6):1397–1409.

Jaworek-Korjakowska, J. and Tadeusiewicz, R. (2013). Hair removal from dermoscopic color images. *Bio-Algorithms and Med-Systems*, 9(2):53–58.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331.

Kilcline, C. and Frieden, I. J. (2008). Infantile hemangiomas: How common are they? a systematic review of the medical literature. *Pediatric Dermatology*, 25(2):168–173.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Land, E. H. and McCann, J. J. (1971). Lightness and retinex theory. *Josa*, 61(1):1–11.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Liang, M. G. and Frieden, I. J. (2014). Infantile and congenital hemangiomas. In *Seminars in pediatric surgery*, volume 23, pages 162–167. Elsevier.

Liao, H. (2016). A deep learning approach to universal skin disease classification. *University of Rochester Department of Computer Science, CSC*.

Lopez, A. R., Giro-i Nieto, X., Burdick, J., and Marques, O. (2017). Skin lesion classification from dermoscopic images using deep learning techniques. In *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*, pages 49–54. IEEE.

Luu, M. and Frieden, I. (2013). Haemangioma: clinical course, complications and management. *British Journal of Dermatology*, 169(1):20–30.

Mattassi, R., Loose, D. A., and Vaghi, M. (2009). *Hemangiomas and Vascular Malformations: An Atlas of Diagnosis and Treatment.* Springer.

Menegola, A., Fornaciali, M., Pires, R., Bittencourt, F. V., Avila, S., and Valle, E. (2017). Knowledge transfer for melanoma screening with deep learning. *arXiv preprint arXiv:1703.07479*.

Mulliken, J. B. and Glowacki, J. (1982). Hemangiomas and vascular malformations in infants and children: a classification based on endothelial characteristics. *Plastic and reconstructive surgery*, 69(3):412–420.

Neghină, C., Zamfir, M., Sultana, A., Ovreiu, E., and Ciuc, M. (2016). Automatic detection of hemangiomas using unsupervised segmentation of regions of interest. In *Communications (COMM), 2016 International Conference on*, pages 69–72. IEEE.

Neghină, C., Zamfir, M., Ciuc, M., and Sultana, A. (2016). Automatic detection of hemangioma through a cascade of self-organizing map clustering and morphological operators. *Procedia Computer Science*, 90:145 – 150. 20th Conference on Medical Image Understanding and Analysis (MIUA 2016).

Oliveira, R. B., Mercedes Filho, E., Ma, Z., Papa, J. P., Pereira, A. S., and Tavares, J. M. R. (2016). Computational methods for the image segmentation of pigmented skin lesions: A review. *Computer methods and programs in biomedicine*, 131:127–141.

Oprisescu, S., Ciuc, M., Sultana, A., and Vasile, I. (2015). Automatic segmentation of infantile hemangiomas within an optimally chosen color space. In *E-Health and Bioengineering Conference (EHB), 2015*, pages 1–4. IEEE.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Püttgen, K. B. (2014). Diagnosis and management of infantile hemangiomas. *Pediatric Clinics of North America*, 61(2):383 – 402. Pediatric Dermatology.

Qi, J., Le, M., Li, C., and Zhou, P. (2017). Global and local information based deep network for skin lesion segmentation. *arXiv preprint arXiv:1703.05467*.

Robertson, A. R. (1977). The cie 1976 color-difference formulae. *Color Research & Application*, 2(1):7–11.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.

Scharcanski, J. and Celebi, M. E. (2014). *Computer vision techniques for the diagnosis of skin cancer*. Springer.

Silveira, M., Nascimento, J. C., Marques, J. S., Marçal, A. R., Mendonça, T., Yamauchi, S., Maeda, J., and Rozeira, J. (2009). Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):35–45.

Sultana, A., Oprisescu, S., and Ciuc, M. (2015a). Automatic evaluation of hemangiomas for follow-up monitoring. In *E-Health and Bioengineering Conference (EHB), 2015*, pages 1–4. IEEE.

Sultana, A., Zamfir, M., Ciuc, M., Oprişescu, Ş., and Popescu, M. (2015b). Automatic segmentation of infantile hemangiomas. In *Signals, Circuits and Systems (ISSCS), 2015 International Symposium on*, pages 1–4. IEEE.

Tollefson, M. M. and Frieden, I. J. (2012). Early growth of infantile hemangiomas: what parents' photographs tell us. *Pediatrics*, 130(2):e314–e320.

Van De Weijer, J., Gevers, T., and Gijsenij, A. (2007). Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214.

Vasconcelos, M. J. M., Rosado, L., and Ferreira, M. (2014). Principal axes-based asymmetry assessment methodology for skin lesion image analysis. In *International Symposium on Visual Computing*, pages 21–31. Springer.

Vogt, M. and Ermert, H. (2006). Multimodal imaging—what can we expect? In *Bioengineering of the Skin: Skin Imaging & Analysis*, pages 17–29. CRC Press.

Xu, C. and Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE Transactions on image processing*, 7(3):359–369.

Zambanini, S., Langs, G., Sablatnig, R., Donath, P., and Maier, H. (2006). Automatic surveying of cutaneous hemangiomas. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 1, pages 1022–1025. IEEE.

Zambanini, S., Langs, G., Sablatnig, R., and Maier, H. (2007). Automatic robust registration of cutaneous hemangiomas for follow-up examinations 1).

Zambanini, S., Sablatnig, R., Maier, H., and Langs, G. (2010). Automatic image-based assessment of lesion development during hemangioma follow-up examinations. *Artificial intelligence in medicine*, 50(2):83–94.

Zhou, H., Schaefer, G., Celebi, M. E., Iyatomi, H., Norton, K.-A., Liu, T., and Lin, F. (2010). Skin lesion segmentation using an improved snake model. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 1974–1977. IEEE.